A Project Based Seminar Report

on

"Machine Learning Algorithms for Cryptocurrency Price Prediction"

Submitted to the

Savitribai Phule Pune University

In partial fulfillment for the award of the Degree of

Bachelor of Engineering

in

Information Technology

by

Himanshu Agrawal

(T150238501 / 307002 & Div 1)

Under the guidance of

Prof. S. M. Jaybhaye

Sinhgad Institutes

Department of Information Technology

STES's Sinhgad College of Engineering

Vadgaon (Bk.), Off. Sinhgad Road,

Pune 411041.

Semester-VI, Third Year Engineering

2017-2018

Sinhgad Institutes

CERTIFICATE

This is to certify that the project based seminar report entitled "Machine Learning Algorithms for Cryptocurrency Price Prediction" being submitted by Himanshu Agrawal (T150238501/307002 and Div 1) is a record of bonafide work carried out by him under the supervision and guidance of Prof. S. M. Jaybhaye in partial fulfillment of the requirement for **TE (Information Technology Engineering) - 2015 Course** of Savitribai Phule Pune University, Pune in the academic year 2017-2018.

Date: 31/03/2018

Place: Pune

Prof. S. M. Jaybhaye
Seminar Guide

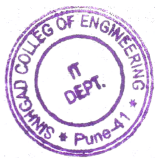
Prof. G. R. Pathak
Head of the Department

Dr. S. D. Lokhande
Principal

This Project Based Seminar report has been examined by us as per the Savitribai Phule Pune University, Pune requirements at STES's Sinhgad College of Engineering, Pune-411041 on . . . . . .

Internal Examiner

External Examiner



ACKNOWLEDGMENT

I am highly indebted to my guide Prof S. M. Jaybhaye for her guidance and constant supervision as well as for providing necessary information regarding the seminar report and also for her support in completing the seminar report. I would like to express my special gratitude and thanks to Staff Members of department of Information Technology for giving me such attention and time.

This acknowledgment would be incomplete without expressing my thanks to Prof. G. R. Pathak, Head of the Department (Information Technology) for his support during the work.

I would like to extend my heartfelt gratitude to my Principal, Dr. S. D. Lokhande who provided a lot of valuable support, mostly being behind the veils of college bureaucracy.

I would also like to express my gratitude towards my parents and friends for their kind co-operation and encouragement which help me in completion of this report. My thanks and appreciations also go to my colleague in developing the seminar report and people who have willingly helped me out with their abilities.

Himanshu Agrawal

ABSTRACT

In recent years,Bitcoin ecosystem has gained the attention of consumers, businesses, investors and speculators alike.As a result of blockchain-network-based feature engineering ,macro-economic factors of actual market and machine learning algorithms optimization, we can obtain up-down Bitcoin price movement classification accuracy of roughly 55 percent. This research is concerned with predicting the price of Bitcoin using machine learning. The goal is to ascertain with what accuracy can the direction of Bit- coin price in USD can be predicted. The price data is sourced from the Bitcoin 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.

CONTENTS

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION TO PROJECT TOPIC

1.1 Introduction to Project

Prediction of mature financial markets such as the stock market has been researched at length. Bitcoin presents an interesting parallel to this as it is a time series prediction problem in a market still in its transient stage. As a result, there is high volatility in the market and this provides an opportunity in terms of prediction. In addition, Bitcoin is the leading cryptocurrency in the world with adoption growing consistently over time. Due to the open nature of Bitcoin it also poses another paradigm as opposed to traditional financial markets. It operates on a decentralised, peer-to-peer and trustless system in which all transactions are posted to an open ledger called the Blockchain. This type of transparency is unheard of in other financial markets. Given the complexity of the task, deep learning makes for an interesting technological solution based on its performance in similar areas. Tasks such as natural language processing which are also sequential in nature and have shown promising results. This type of task uses data of a sequential nature and as a result is similar to a price prediction task. The recurrent neural network (RNN) and the long short term memory (LSTM) flavour of artificial neural networks are favoured over the traditional multilayer perceptron (MLP) due to the temporal nature of the more advanced algorithms.

Various studies on statistical or economical properties and characterizations of Bitcoin prices refer to its capabilities as a financial asset, the relationship between Bitcoin and search information, such as Google Trends and Wikipedia, and wavelet analysis of Bitcoin and BGT-Blockchain Google trends.We here will be bridging the gap between these two by Machine Learning.

1.2 Motivation behind project topic

If you were to pick the three most ridiculous fads of 2017, they would definitely be fidget spinners , artificial intelligence and, yes, cryptocurrencies. I m actually not a hodler of any cryptos. So, while I may not have a ticket to the moon, I can at least get on board the hype train by successfully predicting the price of cryptos by harnessing deep learning, machine learning and artificial intelligence (yes, all of them!). Machine Learning - A computer is said to learn if its performance P on a class of tasks T improves with experience E.Having n number of applications like Image recognition,Sentiment analysis,Natural Language processing,Captcha,Pred analysis and even driverless cars,this field/domain totally amuses one to try something in it. Availability of built in libraries like WEKA to code in JAVA,SCIKITS LEARN in Python,variety of datasets available on kaggle,user friendly editors like Jupyter Notebook and IDE?s like Spyder/Pycharm it become very easy for one to implement his/her ideas.

Bitcoin as a currency is in a transient stage and as a result is considerably more volatile than other currencies such as the USD. Interestingly, it is the top perform- ing currency four out of the last five years. Thus, its prediction offers great potential and this provides motivation for research in the area.

1.3 Aim and Objective(s) of the work

The aim of this project is find out with what accuracy the direction of the price of Bitcoin can be predicted using machine learning methods.

1. Proper data collection and finding out the perfect number and types of variables that can be the best suit.

2. To chose the most appropriate algorithm and ways to validate it,optimize it,evaluate it on various parameters for accuracy and errors.

3. Technology used also played an important role in deciding the success of a project.To select and have hands on these required stuff.

4. Understanding the algorithm used at its full depth so that we can optimize it,generalize it in future.

The purpose of this study is to achieve a considerable amount of accuracy and use all the proposed research to find out the best algorithm for cryptocurrency price prediction.

1.4 Introduction to Seminar Topic

The ubiquity of Internet access has triggered the emergence of currencies. Bitcoin has recently attracted considerable attention.Decentralized,no double spending,security by cryptography/hash/Digital signatures,public ledger,etc have gained the trust of huge crowd.Its journey from quarter of a cent to thousands of dollars have totally fascinated the world.From 21 Million of total,16.7 Million of BTC have already been found.At present 12.5 BTC for new block generation is given and in this way 1800 BTC are generated daily.This whole money math amuses me everytime.

The big AI dream of the world of creating machines as equivalent to human brain is totally possible.Having a lot of interest in all these kind of stuff made me to take my seminar topic as Machine learning algorithms for cryptocurrency price prediction.The other topics of the Project includes Bitcoin and Blockchain,Data extraction,Machine learning. Machine Learning being such a vast domain having applications like Image recognition,Sentiment analysis,NLP,Captcha, Pred analysis and even driverless cars.Wide variety of algorithms to try. Availability of libraries with ready to use function for algo implementation and platforms to compare the accuracies and errors,this may be the concrete reason for me choosing topic as ML prediction Algorithms.

Chapter 2

LITERATURE SURVEY

Different studies have tried to explain various aspects of the Bitcoin such as its price formation, price fluctuations, systems dynamics and economic value.After reading a lot of research papers we came to know what all are the factors that can affect the price of BTC and how can we use them in our Machine learning algorithms to predict the price of BTC.

Reference[1] is the very famous paper which proposed the concept of Bitcoin based on Blockchain technology by S.Nakamoto.The paper proposed the technology behind BTC. Reference[2] being an IEEE research paper was based on Empirical study of a prediction model for cryptocurrency price prediction which uses Bayesian Neural network and Resampling for Model validation.It was found that BNN takes into consideration the non-linear influences as well. From Reference[3],we got to know the standard Market condition i.e Barro's model and the formulas for demand/supply of BTC. Reference[4] gave idea of a perfect market condition.

Reference[5] concludes that BTC price can also be predicted using user comment and replies over Online Crypto Communities.For this Web-Crawler can be used to do sentiment analysis.VADER algorithm analyses user posts, replies,comments,views and other activities on online communities. Furthermore, Reference[6] was a Stanford Study to predict BTC price using Transaction graph.It used Linear regression for a many to one feature mapping and Recurent Neural network.We came to know that www.quandl.com platform has an API to interact with the site www.blockchain.info.So this sort of API's can be used to interact with different websites and collect information for our prediction model. Reference[7] was yet another way to predict BTC price,it uses Twitter sentiment analysis for predicting cryptocurrency price fluctuations.It again used VADER and naive algorithms.So came to know that even social media like Twitter have correlation with BTC price and ML was the way to bridge the gap. Reference[8] was a Thesis that focus on study

of BTC price volatility.It was found that Google trends and the economy indices like SandP 500,VIX also effects BTC price and by Granger-Causality we should study them as well for better results.So Market ups/downs,Google trends,Google searches granger causes BTC price fluctuation. Reference[9] and reference[10] were yet another topmost research papers in the same domain. Reference[11][12][13] were the utmost important links for tracking the BTC price movement at each time.

Chapter 3

ML,BITCOIN AND BLOCKCHAIN

3.1 Machine Learning

Data mining can be defined as the extraction of implicit, previously unknown and potentially useful information from data. Machine learning provides the technical basis for data mining.

The purpose of this research is to predict the direction of the price of Bitcoin. As this is a task with a known target it is a supervised machine learning task although some pre-processing can take advantage of unsupervised learning methods. The super- vised algorithms explored include Wavelets and the wavelet discrete transform, several type of artificial neural networks including the Multi-Layer perceptron (MLP), Elman Recurrent Neural Network (RNN) and Long Short Term Memory (LSTM). In terms of pre-processing, random forests were used for feature selection while Bayesian optimisation was performed to optimize some the parameters of the LSTM.

3.2 Bitcoin

Bitcoin is a digital cryptocurrency and payment system that is entirely decentralized, meaning it is based on peertopeer transactions with no bureaucratic oversight. Transactions and liquidity within the network are instead based on cryptography. The system first emerged formally in 2009 and is currently a thriving open-source community and payment network. Based on the uniqueness of Bitcoin payment protocol and its growing adoption, the Bitcoin ecosystem is gaining lots of attention from businesses, consumers, and investors alike.

3.3 Blockchain

Decentralization can be specified by the following goals: (i) Who will maintain and manage the transaction ledger? (ii) Who will have the right to validate transactions? (iii) Who will create new Bitcoins? The blockchain is the only available technology that can simultaneously achieve these three goals. Generation of blocks in the Blockchain, which is directly involved in the creation and trading of Bitcoins, directly influence the supply and demand of Bitcoins. Combination of Blockchain technologies and the Bitcoin market is a real-world example of a combination of high-level cryptography and market economies.

3.4 Multilayer Perceptron

Simple feed forward neural networks are known as multilayer perceptrons and they form the basis for other neural network models.

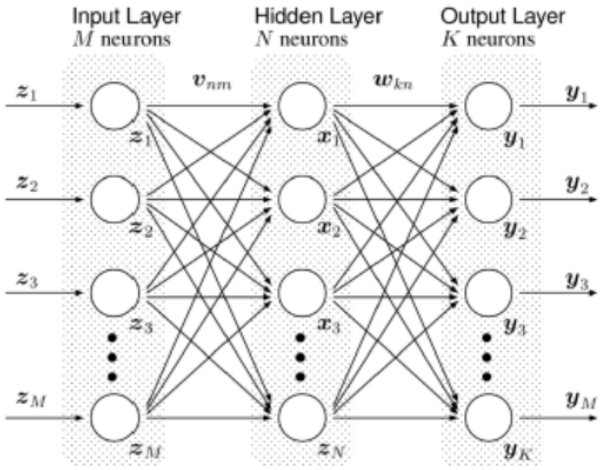


Figure 3.1: Multi-Layer Perceptron

In terms of neural network terminology, examples fed to the model are known as inputs and the predictions are known as outputs. Each modular sub function is a layer. A model consists of an input and output layer, with layers between these known as hidden layers. Each output of one of these layers is a unit which can be considered analogous to a neuron in the brain. The connections between these units is known as a weight which is analogous to a synapse in the brain. The weights define the function of the model as they are the parameter that is adjusted when training a model.

Chapter 4

FACTORS AFFECTING THE PRICE OF BTC.

4.1 Barro's model and BTC economics

Barro's model provides a simple Bitcoin pricing model under perfect market conditions. In this model, Bitcoin is assumed to possess currency value and is exchangeable with traditional currencies, which are under central bank control and can be used for purchasing goods and services. The total Bitcoin supply, SB, is represented by

$$S_B = P_B B$$

where PB denotes the exchange rate between Bitcoin and dollar (i.e. dollar per unit of Bitcoin), and B is the total capacity of Bitcoins in circulation..

The total Bitcoin demand depends on the general price level of goods or services, P; the economy size of Bitcoin, E; and the velocity of Bitcoin, V, which is the frequency at which a unit of Bitcoin is used for purchasing goods or services. The total demand of Bitcoin, DB, is described as followed by:

$$D_B = \frac{PE}{V}$$

The market equilibrium with the perfect market assumption is acquired when the supply and the demand of Bitcoin is the same amount. The equilibrium is therefore achieved at

$$P_B = \frac{PE}{VB}$$

This equilibrium equation implies that in the perfect market, the Bitcoin price in dollars is affected proportionally by the general price level of goods or services multiplied by the economy size of Bitcoin,which can be determined indirectly from the global macroeconomic indexes in actual markets and inversely by the velocity of Bitcoin multiplied by the capacity of the Bitcoin market,extracted from the Blockchain platform.

4.2 Types of Variables

There are mostly three types of variables namely,Response variables which includes log price of BTC,Blockchain information like block size,transactions per block,confirmation time,hash/difficulty rate,macro-ecnomic factors consisting of SandP500, DOW30,NASDAQ,CrudeOil,Gold,Nikkei225,VIX and Global currnecy ratio i.e IPY,CHE,CNY,EUR.

4.3 Dependent and Independent Variables

The independent variable for this study is the closing price of Bitcoin in US Dollars taken from the Coindesk Bitcoin Price Index. Rather than focusing on one specific exchange this price index takes the average prices from five major Bitcoin exchanges; Bitstamp, Bitfinex, Coinbase, OkCoin and itBit. The closing price is chosen over a three-class dummy classification variable representing price going up, down or staying the same for the following reason; the use of a regression model over a classification model offers further model comparison potential through the capture of the root mean squared error (RMSE) of the models. Classifications are then made based on the prediction of the regression model e.g. price up, price down or no change. Additional performance metrics include accuracy, specificity, sensitivity and precision.

The dependent variables are taken from the Coindesk website, Blockchain.info and from the process of feature engineering.

Chapter 5

MACHINE LEARNING ALGORITHMS.

5.1 Recurrent Neural network vs Long Short Term memory Algorithm

The recurrent neural network (RNN) was first developed by Elman. The

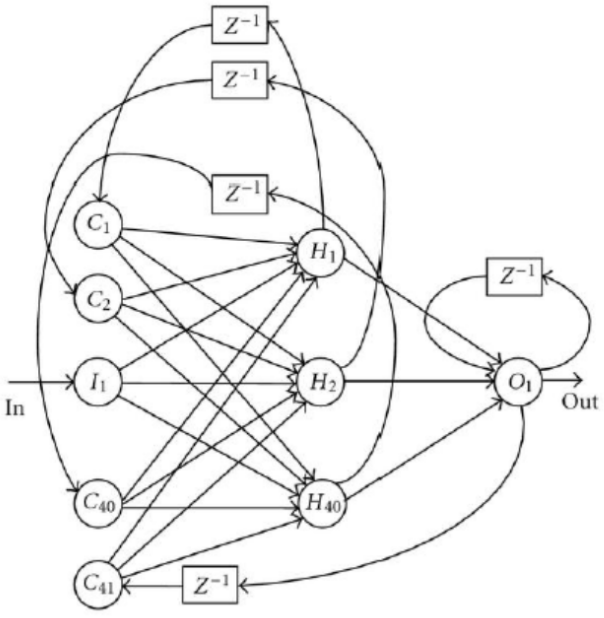


Figure 5.1: Recurrent Neural Network

RNN is structured similarly to the MLP, with the exception that signals can flow both forward and backwards in an iterative manner. To facilitate this

another layer known as the context layer is added. In addition to passing input between layers, the output of each layer is fed to the context layer to be fed into the next layer with the next input.

In this context, the state is overwritten at each timestep. This offers the benefit of allowing the network to assign particular weights to events that occur in a series rather than the same weights to all input as with the MLP. This results in a dynamic network.

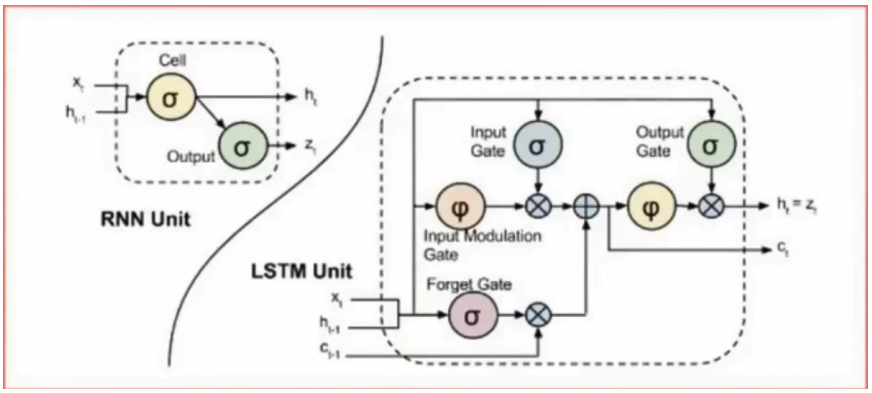


Figure 5.2: RNN vs LSTM

The Long Short-Term Memory network, or LSTM network developed by Hochreiter et al.,they allow the preservation of the weights that are forward and back-propagated through layers. This is in contrast to the Elman RNN in which the state gets overwritten at each step. They also allow the network to continue to learn over many time steps by maintaining a more constant error. This allows the network to learn long term dependencies.It is a recurrent neural network that is trained using Backpropagation Through Time and overcomes the vanishing gradient problem.As such, it can be used to create large recurrent networks that in turn can be used to address difficult sequence problems in machine learning and achieve state-of-the-art results.

Instead of neurons, LSTM networks have memory blocks that are connected through layers.A block has components that make it smarter than a classical neuron and a memory for recent sequences. A block contains gates that manage the block's state and output. A block operates upon an input sequence and each gate within a block uses the sigmoid activation units to control whether they are triggered or not, making the change of state and addition of information flowing through the block conditional. Each unit is like a mini-state machine where the gates of the units have weights that are learned during the training procedure.

5.2 Bayesian Neural Network

Bayesian neural networks (BNN) is a transformed Multilayer perceptron (MLP) which is a general term for ANNs in the fields of machine learning. The networks have been successful in many application such as image recognition, pattern recognition, natural language processing, and financial time series. It becomes known that much effective to represent the complex time series than the conventional linear models, i.e. autoregressive and moving average, etc. The structure of a BNN is constructed with a number of processing units classified into three categories: an input layer, an output layer, and one or more hidden layers.Different from a single layer perceptron, which can only be linearly separated, they solve XOR problems by introducing backpropagation algorithms and hidden layers. The hidden layer mapping the original data to a new space transforms data that cannot be linearly separated into linearly separable data.

Weights of a BNN must be learned between the input-hidden layer and hidden-output layer. Backpropagation refers to the process in which weights of hidden layers are adjusted by the error of hidden layers propagated by the error of the output layer. An optimization method called delta rule is used to minimize the difference between a target value and output value when deriving backpropagation algorithm. In general, BNNs minimize the sum of the following errors, EB, using backpropagation algorithm and delta rule.

$$E_B = \frac{\alpha}{2}\sum_{n=1}^{N}\sum_{k=1}^{K}(t_{nk} - o_{nk})^2 + \frac{\beta}{2}\precapprox_B^T \precapprox_B$$

where EB is the sum of the errors, N is the number of the training variables, K is the size of the output layer, tnk is the kth variable of the nth target vector, onk is the kth output variable of the nth training vector,alpha and beta are the hyperparameter, and wB is the weights vector of the Bayesian neural network.

5.3 Resampling by Bootstrap and Cross Validation

Resampling refers to a method used for model validation.We have two-cross-validation, and bootstrap.We identify advantages and disadvantages of each method and select the appropriate method for the empirical analysis of this study.

A bootstrap method is one of the sampling techniques that new data set is sampled from the original data set with the replacement. A typical bootstrap works as follows-

1 . We have the original data setD with the number of N.

2 . Below following step is repeated B times for particular large number to produce B different bootstrap data set, Z1; Z2; . . . ; ZB here, Data set Zi with the size N is generated by sampling from the original data set D with the replacement.

3 . The machine is trained from each bootstrap data set.

4 . Accuracy of the machine is calculated by averaging each bootstrap data set.

$$Accuracy = \frac{1}{B}\sum_{j=1}^{B}\frac{1}{N}\sum_{i=1}^{N}(1 - Loss(\hat{y}_i^j, y_i))$$

where yi is an i-th true training output data, yij i is an i-th estimated output from the bootstrap data Zj, and Loss() is a loss function.

A cross-validation randomly divides the original data set into K equal-sized parts without the replacement. We fit the machine learning model to the K - 1 parts leaving out particular set k and acquire a prediction error for the left-out k part. Total prediction accuracy is combined after the procedure is repeated for each part to leave. A general procedure is as follows:

1 . We divide the original data set into K partial equal-sized data.

2 . We can compute the total accuracy:

$$accuracy_K = \sum_{k=1}^{K}\frac{n_k}{N}\frac{1}{n_k}\sum_{i=1}^{n_k}(1 - Loss(\hat{y}_i^k, y_i))$$

where N is the total number of the original data set, others have same definition with in the bootstrap description.

3 . The estimated standard deviation of the cross-validation:

$$\hat{SE}(CV_K) = \sqrt{\frac{\sum_{k=1}^{K}(Err_k - \bar{Err_k})^2}{N-1}}$$

Err_k is the k-th loss, $\sum_{i=1}^{n_k} Loss(\hat{y}_i^k, y_i)$.

Bootstrap is adequate to validate a predictive model performance, to use an ensemble method, and to estimate of bias and variance of the trained model. Bootstrap creating the cloned multiple samples with the replacement is not originally developed for model validation. It can give more biased results. Therefore, we employ the cross-validation technique to our model validation. Cross-validation can create high-variance problems when data size is small. Our data size is sufficient to overcome the problem. We employ the 10-fold cross-validation methods generally used for model validations.

5.4 Statistical Analysis of Algorithms

Table 5.1: Statistical Analysis of Algorithms

Response variable	**Log Price-RMSE**	**Log Volatility-RMSE**
Linear regression	0.0935	0.4823
Bayesian NN	0.0039	0.2325
Support Vector regression	0.3201	0.5297

The Root Mean Square Error of three different algoithms can be analysed from the above table.Bayesian NN having the least RMSE for log price and log volatility,it can be chosen as the best amongst.

5.5 Architecture and Block Diagram

Prediction model for time series analysis usually consists of two phases namely **1) Model Creation** and **2) Using the Model with Streaming data**. The first phase is to built a model that can be deployed.For this data set which is time series analysis is used.Various factors collected from different sites through API are then passed through feature selection and feature extraction process,and this is where we get our final data set.Further train and test splits is done.About 70 percent of data is train set and remaining 30 percent is for test set.

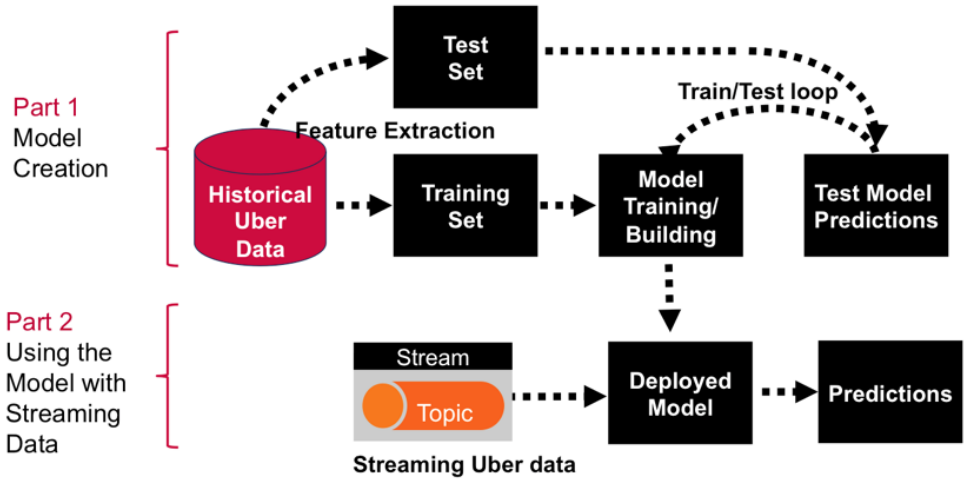


Figure 5.3: Prediction Model architecture

There is a slight line between overfitting and generalization which becomes a matter of concern.If we try to overfit the model then it wont be able to generalize in future and since our time series is streaming continuously,generalization is must.The train set is fed to ML algo and finally a model is built which can then give accurate results for the test set.Test data is usually fed in streaming way. 10 fold cross validation of model is done which is a loop of train and test over different segments of data set and then taking average.This is for better performance.

Finally,Accuracy of prediction is calculated by giving new test data. Since we will be predicting one hour prior so results and calculated perse. The final deployed model is then ready for prediction.

5.6 Structure of the Experiment

Most of the previous studies have focused on either modeling Bitcoin price without considering its relationship to Blockchain information or identifying only its "linear" relationship to macroeconomic factors. The present study attempts to overcome these limitations by employing a Bayesian NN model that can investigate nonlinear influences of each relevant feature of input variables, the Blockchain information, and macroeconomic factors, on Bitcoin price formation. To this end, we first train a Bayesian NN to model Bitcoin price formation using given above-mentioned relevant features of the process.We have evaluated Bayesian NN in terms of training and test errors by using the representative non-linear methodologies, SVR, and the linear regression model as the benchmark methods.

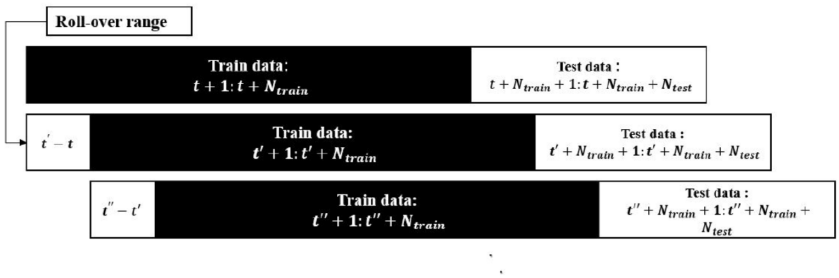


Figure 5.4: Roll-over Framework Strategy

Next, we develop a prediction model of the near-future price of Bitcoin after modeling the entire process. We configure forecasting models by the rollover framework,which is generally applied to portfolio theory.**Rollover strategy** is known as rolling a position forward which is closing out an old position and establishing a new position in a contract of the portfolio with a long time to maturity.

In our experiments, the trained machine is closing out an old information and acquiring new data according to the rollover framework over time,in this way, and the average performance of prediction errors measured several times is evaluated.

Given that the model employs time series in batch format, it is is faster and easier to learn than other sequential neural networks models, LSTM or RNN, and can reflect the flow of information that changes with time.

Chapter 6

TECHNOLOGIES USED.

We will be using Scikit learn library which contain built in standard functions to implement various machine learning algorithms.Jupyter notebook is the editor which comes inbuilt with Anaconda (python distribution).WEKA can be used for comparison between different algorithms.

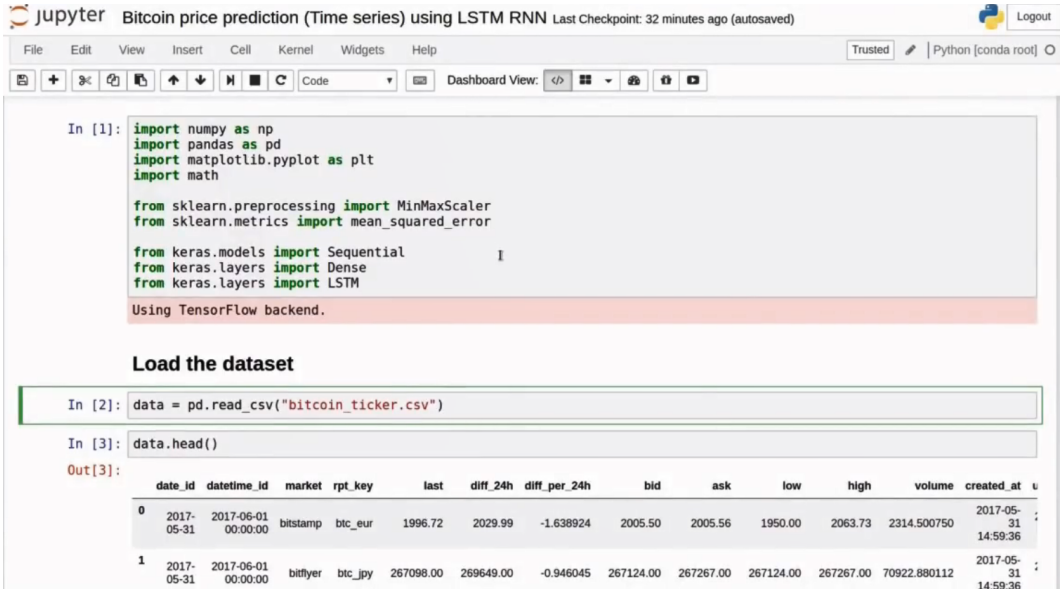


Figure 6.1: Importing datasets and using Standard functions

The snap shows how libraries are loaded and dataset is imported.numpy, pandas, matplotlib are various libraries containing standard functions.

LSTM model is build and compiled.Entire dataset is splitted into train and test.Data filtering,Sequencing according to time,creating numpy arrays, scaling of the data and finally appropriate plots are various tasks done as shown in the snap.

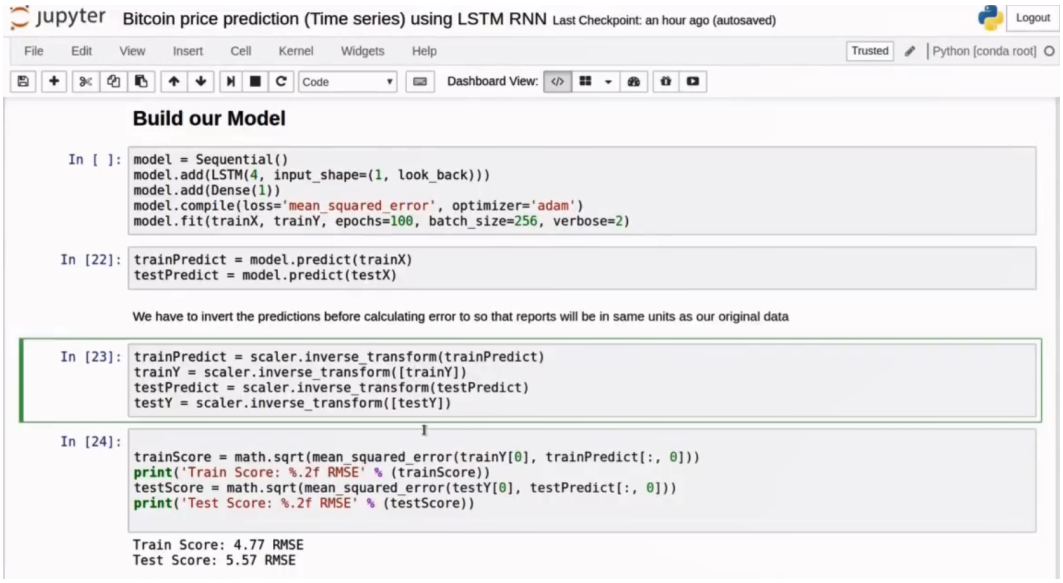


Figure 6.2: Methodology using jupyter nb,Scikit learn

Data collected form various sites through API can be imported here.Python is used to code.MATLAB may also be required for statistical analysis of data.WEKA proves to be very useful to compare the accuracies of various algorithms.Sklearn contains ready to use function for implementing algorithms, for instance KNN algorithm can be used as sklearn.knn().Finally the predict function returns the accuracy and RMSE values.

Chapter 7

PROS/CONS AND APPLICATIONS

7.1 Pros and Cons

The highly volatile nature of BTC pricing can be solved using a proper machine learning approach.Availability of various libraries with standard built in functions like WEKA,SCIKIT learn,Tensorflow proved to be very useful.API's to interact with different sites to obtain data.Lot on research been already done on this very topic came out to be very useful and proved to be the advantageous part.

Accuracy of the proposed model can be a matter of little concern.This was a point that can be disadvantageous.Prediction range at this point can be one hour prior to the actual event.

7.2 Applications of BTC price prediction model

A software/web-page/application based on the proposed research which keep a track on all the factors fetched online/offline time series can prove to very useful by the online crypto-communities to invest BTC.Bussiness organizations,economists can use this model for further development.As the variation of Bitcoin process gets attention, it is expected that the expansion and application of the BNN model would be effective for the analysis and prediction of the Bitcoin process. The unpredictable prices prob solved.

Chapter 8

CONCLUSION , FUTURE WORK AND REFERENCES

8.1 Conclusion

Looking back we can conclude that Deep learning models such as the RNN and LSTM are evidently effective learners on training data with the LSTM more capable for recognising longer-term dependencies. However, a high variance task of this nature make it difficult to transpire this into impressive validation results. As a result it remains a difficult task. There is a fine line to balance between overfitting a model and preventing it from learning sufficiently. The goal can be achieved by adopting other extended machine learning methods or considering new input capabilities related to the variability of Bitcoin. Such study will contribute to rich Bitcoin time series analysis in addition to existing Bitcoin studies. This seminar is undertaken to explain what all factors affect the price of BTC and how we can use them in proper Machine Learning algorithms and to evaluate the accuracy of Prediction.This study has found that ML bridges the gap between Statistics and BTC price.

8.2 Reference

1. S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.
2. H. Jang , J. Lee - "An Empirical Study on Modeling and Prediction of Bitcoin Prices With Bayesian Neural Networks Based on Blockchain Information," December 4, 2017.
3. R. J. Barro, "Money and the price level under the gold standard," Econ. J.,vol. 89, no. 353, pp. 13-33, 1979.
4. P. Ciaian, M. Rajcaniova, and D. Kancs, "The economics of Bitcoin price formation," Appl. Econ., vol. 48, no. 19, pp. 1799-1815, 2016.
5. Y. B. Kim1, Jun Gi Kim2, W. Kim3, Jae Ho Im3, T. H. Kim1, Shin Jin Kang2, Chang Hun Kim3,"Predicting Fluctuations in Cryptocurrency Transactions Based on User Comments and Replies," August 17, 2016.
6. A. Greaves, B. Au,"Using the Bitcoin Transaction Graph to Predict the Price of Bitcoin," December 8, 2015.
7. E. Stenqvist,J. Lonno - "Predicting Bitcoin price fluctuation with Twitter sentiment analysis",Kth Royal Institute Of Technology,Stockholm, Sweden 2017.
8. J. C. Soldevilla Estrada,"Analyzing Bitcoin Price Volatility," University of California, Berkeley ,May 5, 2017.
9. S. McNally, "Predicting the price of Bitcoin using machine learning," Ph.D. dissertation, School Comput., Nat. College Ireland, Dublin, Ireland, 2016.
10. I. Madan, S. Saluja, and A. Zhao, "Automated Bitcoin trading via machine learning algorithms," Dept. Comput. Sci., Stanford Univ., Stanford, CA, USA, Tech. Rep., 2015.
11. Blockchain Info. https://blockchain.info/
12. Coinbase API. https://www.coinbase.com/docs/api/overview
13. OKCoin API. https://www.okcoin.com/about/publicApi.do