Predicting the Bitcoin For Future
How much more could the Cryptocurrency rise?

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Computer Science and Engineering Department

Thapar Institute of Engineering and Technology, Patiala

May 2024

TABLE OF CONTENTS

Project Overview(Abstract)

This project aims to address the challenge of predicting the price & Analysis of change of Bitcoin along the time depending upon the it’s Volume & the circumstances along the World’s Economy by using the Machine Learning Techniques.

By Analysing historical Bitcoin price data along with relevant market indicators, the Model seeks to forecast future price trends accurately.

Machine Learning Algorithms, including Regression, Time Series Analysis, and Deep Learning are employed to build and train the prediction model. Various techniques like- Linear Regression, ARIMA, LSTM(Long Short Term Memory) and ensemble methods are explored and evaluated for their effectiveness in forecasting the prices.

The project involves several key steps. Firstly, comprehensive Data Collection of Historical Bitcoin prices and associated market variables will be conducted. Subsequently, exploratory Data Analysis are performed to gain insights into the underlying patterns and correlations.

The ultimate goal of this project is to develop a reliable Bitcoin Price Prediction model that can assist investors, traders and other stakeholders in making informed decision in the volatile cryptocurrency market. The insights gained from this research can also contribute to the broader understanding of market dynamics and potentially the pave the way for future advancements in cryptocurrencies forecasting methodologies.

Problem Statement

Bitcoin Changing price per minute/ per second.

"The volatile nature and massive potential for return make the Cryptocurrency, especially Bitcoin the investment of interest. The major challenge is predicting the future price of Bitcoin. Developing a robust Bitcoin Price Prediction using advance AI techniques to accurately forecast future price movements in the volatile cryptocurrency market."

Developing robust and adaptable Bitcoin Price Prediction solutions is imperative to mitigate the multifaceted risks posed by changing the the flow of currency along time. These solutions should leverage advanced algorithms and models to detect subtle inconsistencies and variables indicative of Bitcoin Price. Halving events lead to lower supply, with fewer Bitcoins made available, leading to their higher prices. Bitcoin as the world's most valuable cryptocurrency and is traded on over 40 exchanges worldwide accepting over 30 different currencies offers a novel opportunity for price prediction due to its relatively young age and resulting volatility, which is far better than that of flat currencies.

Need Analysis

Scope:

- Develop solution for Bitcoin Volatile Price by making an AI predictor.
- Create a mechanism for processing time series analysis of Bitcoin behaviour over last 12 years.

Motivation:

- Bitcoin presents an interesting parallel to the mature financial market as it is a time series prediction problem in a market still in a transient stage.
- Bitcoin has a current market capitalisation of 9 Billion USD and sees over 250, 000 transactions per day.
- Recognise that with what accuracy the price of Bitcoin can be predicted using Machine Learning and compare parallelism methods execute on GPU environments.
- After the boom and bust of Cryptocurrencies, Bitcoin has been increasingly regarded as an investment asset. Because of its highly volatile nature, there is a need for good prediction on which to base investment decisions.
- The investigation of Bitcoin price prediction will be a pilot study of the importance of sample dimension in Machine Learning techniques.

To facilitate a comparison to more traditional approaches in financial forecasting, an ARIMA time series model is used for performance comparison purpose with neural network model.

The purpose of this model is to obtain an algorithm model with high prediction accuracy for the price of Bitcoin on the next day through random forest regression and LSTM, and to explain which variables have influence on the price of Bitcoin. Prediction accuracy of Random Forest regression is significantly better than that of LSTM. The changes in the variables that determine the price of Bitcoin in each period are also obtained through Random Forest regression. The relationship between accuracy and the number of periods of explanatory variables bought into the model shows that for predicting the price of Bitcoin for the next day, the model with only one lag of the explanatory variables has the best prediction accuracy.

The consequences of applying Deep Learning models such as LSTM and RNN are evidently effective for Bitcoin Price Prediction with the LSTM more capable of recognising longer term dependencies. The ability to predict using streaming data should improve the model. Sliding window validation is an approach that can used for prediction task.

Novelty

While making such Bitcoin Price Prediction model it is not easy to bring novelty because so much for is already done towards this path. But in this model I try to bring some sort novelty by combining the accuracy and results of different Machine learning Models such as LSTM(Long Short Term Memory), Linear Regression, Random Forest Regressor. In addition to to pursuing a high-precision forecasting model, this study also conduct in-depth analysis from the explanatory variables that determine the importance of Bitcoin prices and the relationship between the prediction accuracy and the lag of explanatory variables.

Despite of using Combination of models , I also computes the accuracy with same models/ regressors using Cuda GPU computing foe better computation and having less time complexity.

I mainly focus on the performance of random forest regression in Bitcoin Price Prediction when using the prediction result of LSTM as a comparison. The prediction effect of random forest in predicting stock price direction has been proven effective. However unlike Random Forest classifier, whose research goals is to classify ups or downs, there are not many papers that use random forest regression to study the Cryptocurrency market in the existing literature.

Combining multiple Machine Learning techniques, such as traditional statistical methods with Deep Learning Architectures or ensemble methods leverages the strength of each approach. This hybrid approach can lead to improve accuracy and robustness in predicting Bitcoin Prices. This model also integrate insights from Diverse Fields such as Behavioural economics, network science, or complex system theory into the model development process.

Objectives

We aim to achieve these objectives in this project:

- **To achieve high accuracy:** Develop a model capable of accurately predicting Bitcoin prices within specified time frame, aiming for high-level of accuracy to provide reliable forecasts for investors & traders.
- **To capture Short-term Fluctuations & Long-term Trends:** Design the model to capture both short-term price fluctuations and long-term trends in the Bitcoin-market, allowing stakeholders to make informed decisions for various trading strategies.
- **To Incorporate Relevant Market Indicators:** Integrate a diverse set of market indicators, including historical price data, trading volumes, market sentiment, volatility measures and macroeconomic factors, to enhance the model's predictive capabilities and robustness.
- **To optimise Risk-return Tradeoffs :** Utilise the model to provide decision support for investors and traders offering actionable insights and recommendations based on predicted price movement of Bitcoin.
- **To promote Ethical & Responsible Use:** Ensure that the model is developed and deployed in a manner that upholds ethical principles and promotes responsible use, addressing issues such as bias, fairness, transparency, and privacy in the prediction process.
- **To Enhance Interpretability:** Strive to enhance the interpretability of the model by providing insights into the key factors driving price predictions, enabling stakeholders to understand the rationale behind the forecasts and make informed decisions.

Methodology

The proposed Bitcoin Price Detection system is based on various Machine Learning and Deep Learning Techniques like:- ARIMA, LSTM, RNN, and combination of Regressors.

The accurate system for Bitcoin Price Detection consists of four phases such as:

1. Pre-processing

2. Applying Machine Learning Models

3. Evaluating Accuracy

4. Predicting the final Outcome/Price

Machine learning is an important branch of artificial intelligence (AI). According to whether there is a target variable, it can be divided into supervised learning, unsupervised learning, and reinforcement learning. The purpose of this study is to predict future Bitcoin prices, so a regression function with supervised learning is used. The unified execution logic of machine learning is that after the algorithm is preset, a learner is generated, and a high-precision learner is obtained by repeated training of the learner through training data and the process of validation. Finally, the test data is substituted into the trained learner for evaluation and application.
Both random forest regression and LSTM model training in this paper are implemented through the open-source library of python's machine learning. The library used by random forest regression is sklearn, and LSTM uses keras for research. The pre-processing and collation of the data are done by pandas.

	Timestamp	Open	High	Low	Close	Volume_(BTC)	Volume_(Currency)	Weighted_Price
0	1325317920	4.39	4.39	4.39	4.39	0.455581	2.000000	4.390000
1	1325317980	4.39	4.39	4.39	4.39	0.455581	2.000000	4.390000
2	1325318040	4.39	4.39	4.39	4.39	0.455581	2.000000	4.390000
3	1325318100	4.39	4.39	4.39	4.39	0.455581	2.000000	4.390000
4	1325318160	4.39	4.39	4.39	4.39	0.455581	2.000000	4.390000
...	...	...	...	...	...	...	...	...
855194	1376629560	97.00	97.00	96.81	96.81	58.000000	5621.769821	96.927066
855195	1376629620	97.00	97.00	96.81	96.81	58.000000	5621.769821	96.927066
855196	1376629680	97.00	97.00	96.81	96.81	58.000000	5621.769821	96.927066
855197	1376629740	97.00	97.00	96.81	96.81	58.000000	5621.769821	96.927066
855198	1376629800	97.00	97.00	96.81	96.81	58.000000	5621.769821	96.927066

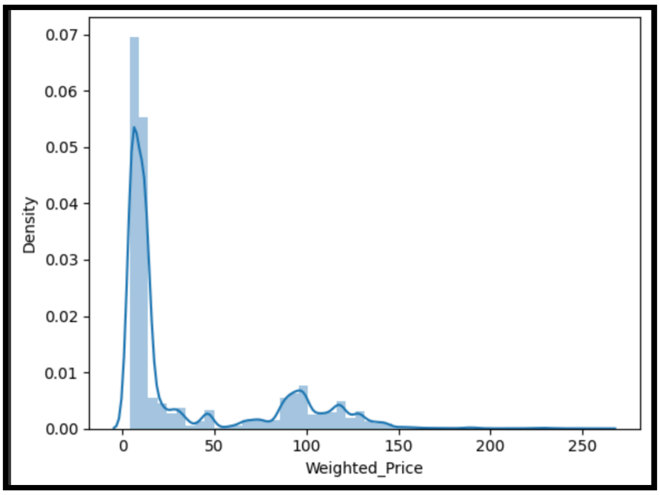
855199 rows × 8 columns

Bitcoin Prediction Dataset

Pre-Processing:

For predicting the Bitcoin Price it is necessary to take the historical data of last10-15 years and to clean & preprocess it first to get a proper sequence of data that how Bitcoin prices will change according to the changing timestamps.

The first step is to gather historical Bitcoin price data. ‘bitstampUSD_1 min_data_2012-01-01_to_2018-03-27.csv' file which contains daily price information.



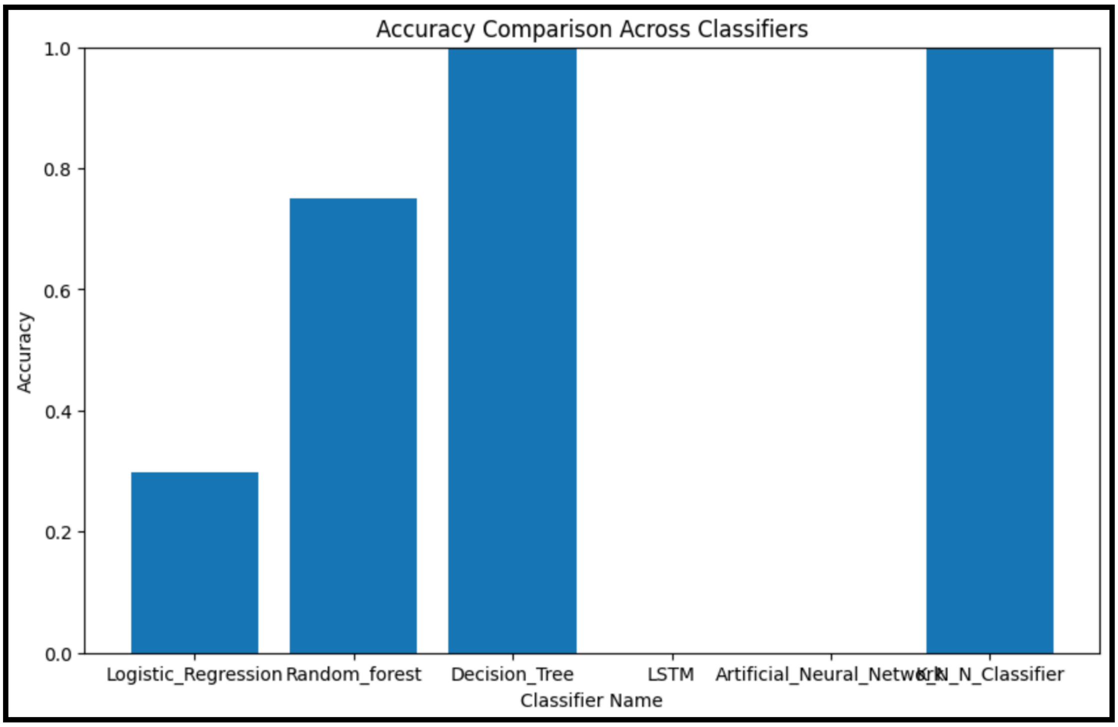
SNS plot of Weighted Price of Bitcoins

	Coefficient
Open	11.105351
High	16.968238
Low	15.091153
Volume_(BTC)	-0.003572

X Values taken from Dataset

Machine Learning Models:

First of all the accuracy of the different Machine Learning models are generated after applying them on the Bitcoin Historical DataSet. These Machine Learning Models involves:- Logistic Regression to divide the values in two categories, Random Forest Classifier to calculate the accuracy of prediction, Support Vector Machine, Decision Tree Classifier, Artificial Neural Network , AdaBoost Classifier and KNN Classifier. After applying all these models , the accuracy of all the models are compared.



Comparison of Different Machine Learning Classifiers according to their Accuracy

```
cuml's r2 score :  1.0
sklearn's r2 score :  0.9999999999999466
['LR.model']
```

Accuracy of CuML's Regression Model(Make Regression)

```
Accuracy for Bitcoin price Prediction 0.3543532302494553
```

Accuracy of AdaBoostClassifier

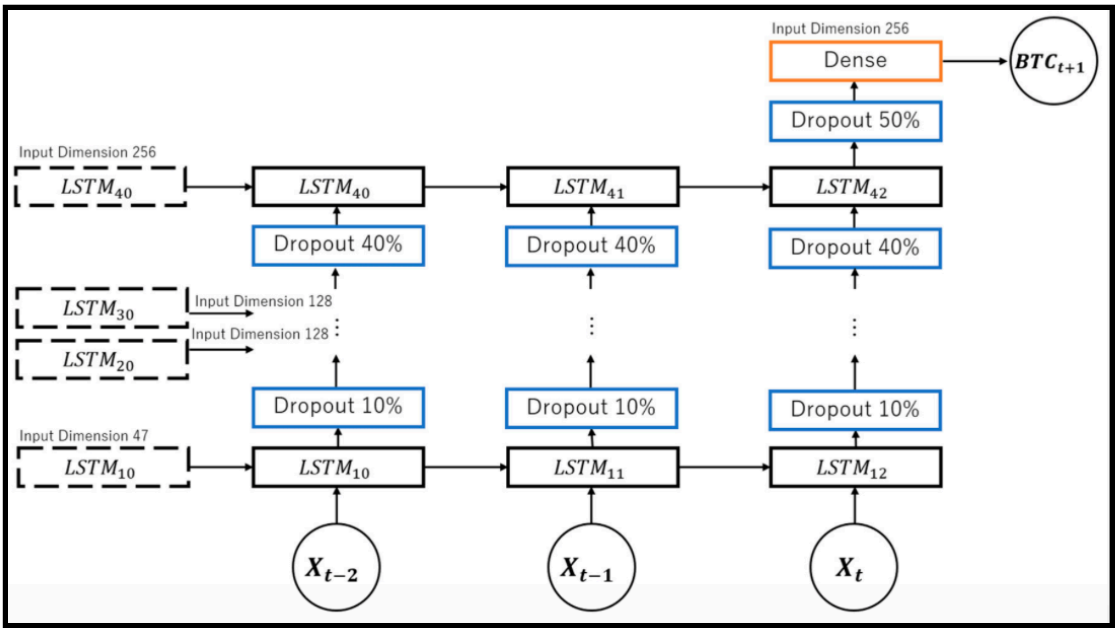
LSTM

When it comes to predicting the price of Bitcoin, the LSTM model has been found to be quite effective. It is a type of recurrent neural network (RNN) that can capture long-term dependencies and patterns in time-series data, which makes it well suited for predicting sequences of values like a time series of Bitcoin prices.

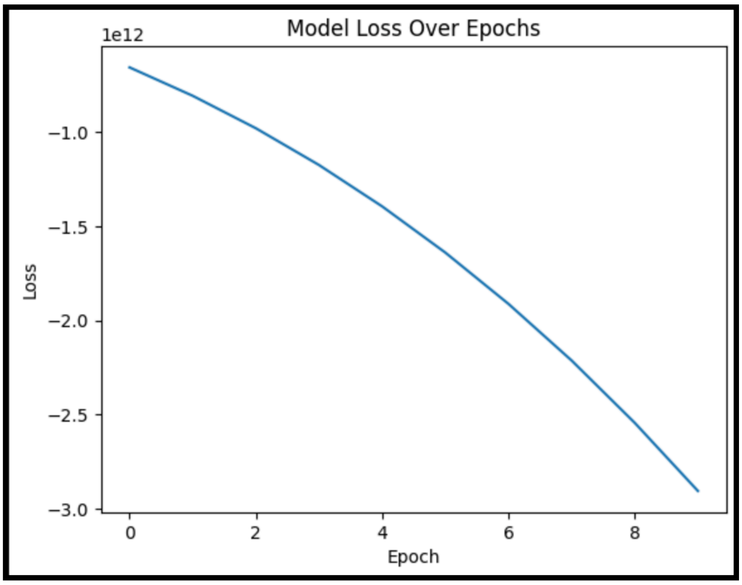
The LSTM model we've built works by taking a sequence of past Bitcoin prices as input and outputting a predicted price. The model is trained on a portion of the data (the training set), and its performance is evaluated on a portion of the data that it hasn't seen during training (the test set). The goal is to minimise the difference (or error) between the model's predictions and the actual prices in the test set.

The LSTM layers in the model are responsible for learning the dependencies between the prices in the input sequence and the output prediction. The Dense layers following the LSTM layers are traditional fully connected neural network layers that generate the final output prediction.

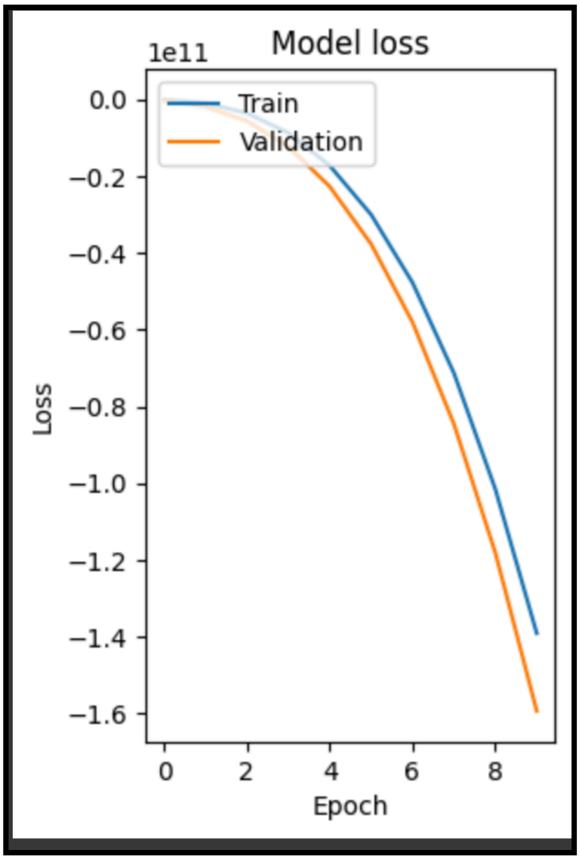
Despite its simplicity, this LSTM model can capture complex patterns in the time-series data, and with enough training data, it can make fairly accurate predictions.



Based on the characteristic that the output value of the LSTM model can be re substituted into another layer of the LSTM model, and the application of the dropout layer mentioned in the literature, the LSTM model structure of this experiment is as follows. Regarding the parameter setting of the dropout layer, I tested [min = 10%, max = 50%] for each dropout layer. It turns out that when the overall value of dropout is small, there is an over learning phenomenon in which the training data performs well but the prediction error of validation data is large. When the overall value of dropout is set too large, the errors of the training data and the validation data are both large. In addition, the experiment also found that the prediction accuracy of the dropout value with descending order is worse than ascending order. The number of layers of LSTM [min = 2, max = 6] and the parameter setting of each layer of units in [32, 64, 128, 256, 512] are tested. After balancing the accuracy and the risk of over learning. The activation function of each layer is set to “ReLU”, which has better performance than “sigmoid” and “tanh”.



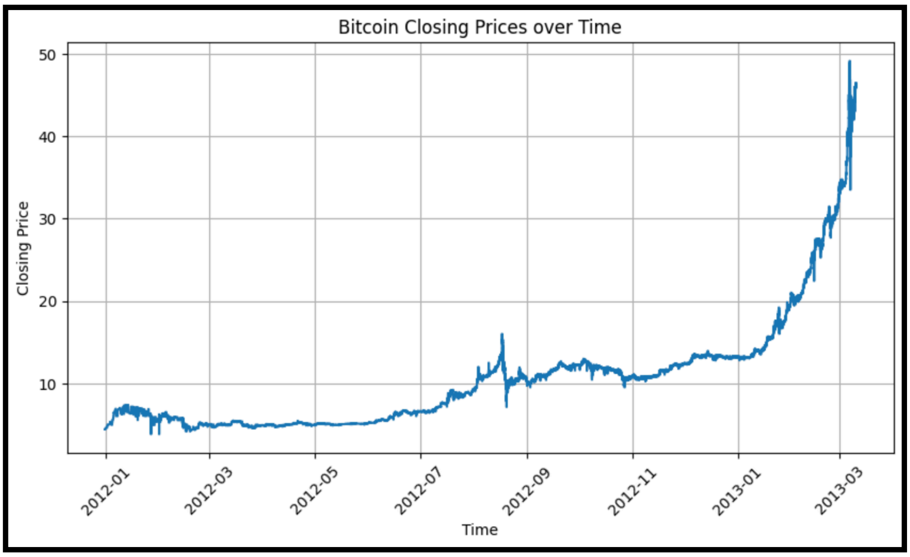
Loss Epochs of LSTM Model



Plot Training and Validation Loss Values

ARIMA (AutoRegressive Integrated Moving Average):

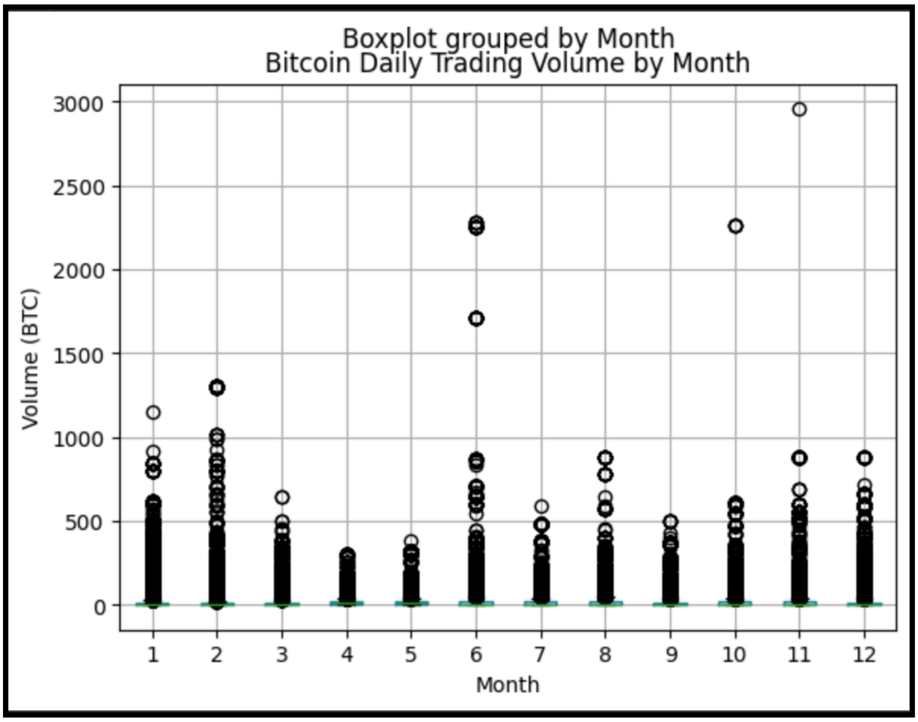
ARIMA is a popular statistical method for time series forecasting. ARIMA models take into account the past values to predict the future values. There are three important parameters in ARIMA: p (past values used for forecasting the next value), q (past forecast errors used to predict the future values), and d (order of differencing).



Plot For Bitcoin Closing Price over time

SARIMA(Seasonal Autoregressive Integrated Moving Average):

SARIMA extends ARIMA by incorporating seasonal components, making it suitable for time series data with seasonal patterns. The seasonal components include additional parameters for seasonal autoregressive, differencing, and moving average terms.

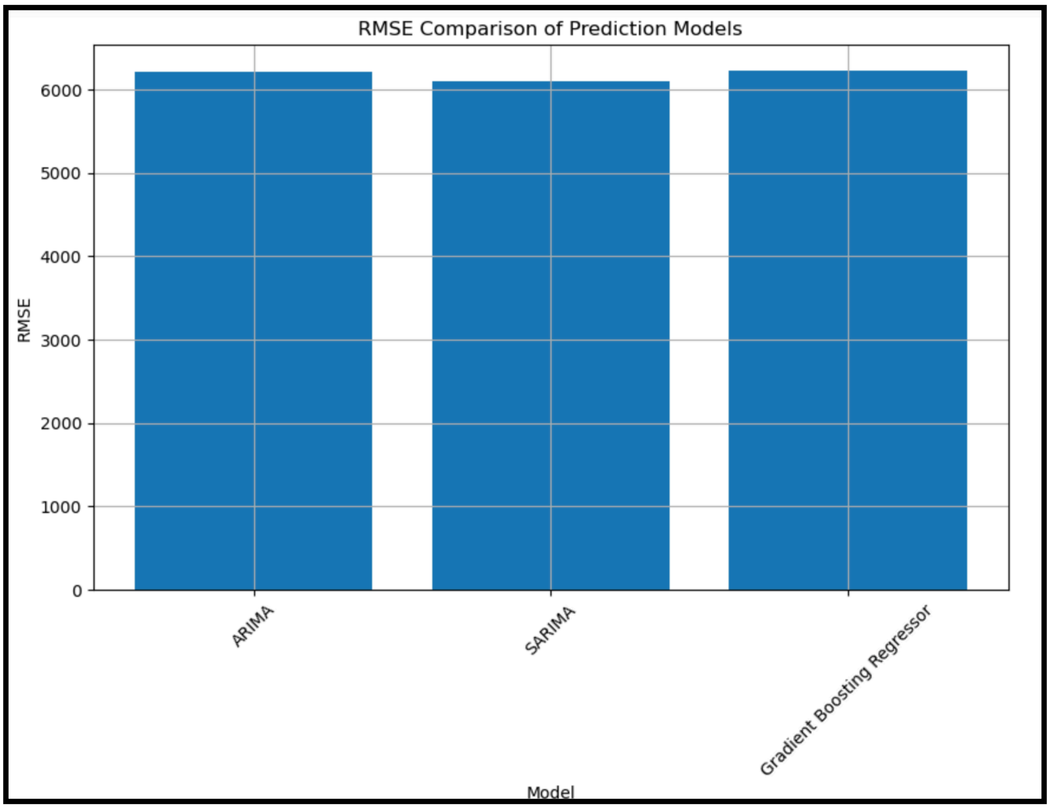


Plot Box of Bitcoin daily trading volume by month

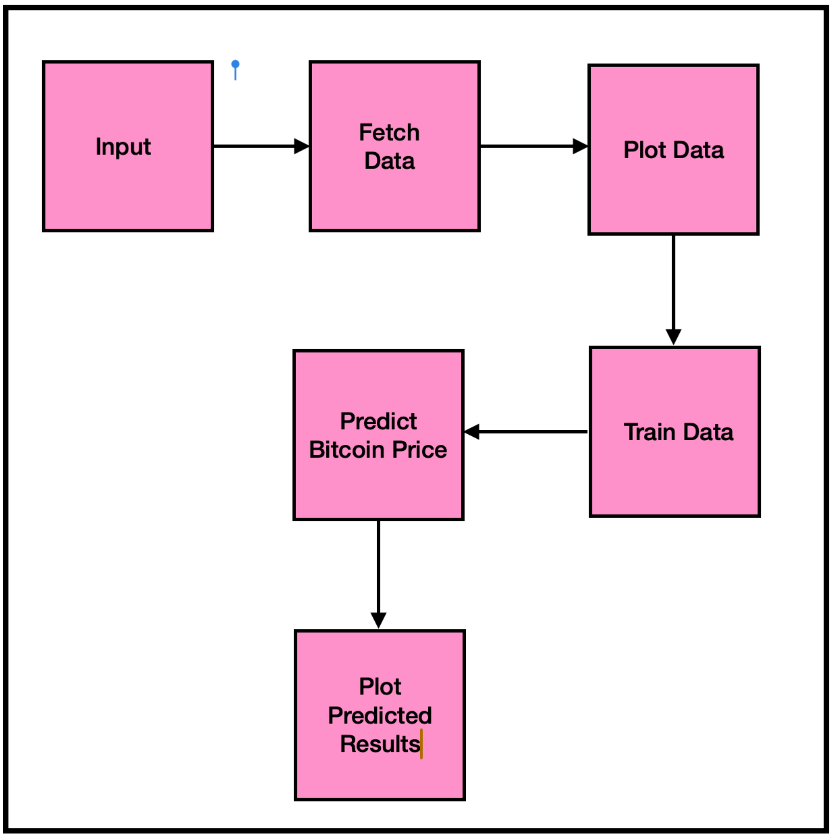
Model Comparison

A confusion matrix representing the ratio of true/false and positive/negative classifications is used to derive the ratings metrics. Accuracy can be defined as the total number of correctly classified predictions (price up, down, and no change). To combat inherent class imbalance (bitcoin price predominately increases) the metrics sensitivity, specificity and precision are also analysed. Sensitivity represents how good a model is at detecting positives. Specificity represent how good the model is at avoiding false alarms. Finally, precision represents how many positively classified predictions were relevant. Root Mean Square Error (RMSE) is used to evaluate and compare the regression accuracy. To instrument the evaluation of models, a 80/20 holdout validation strategy is used.

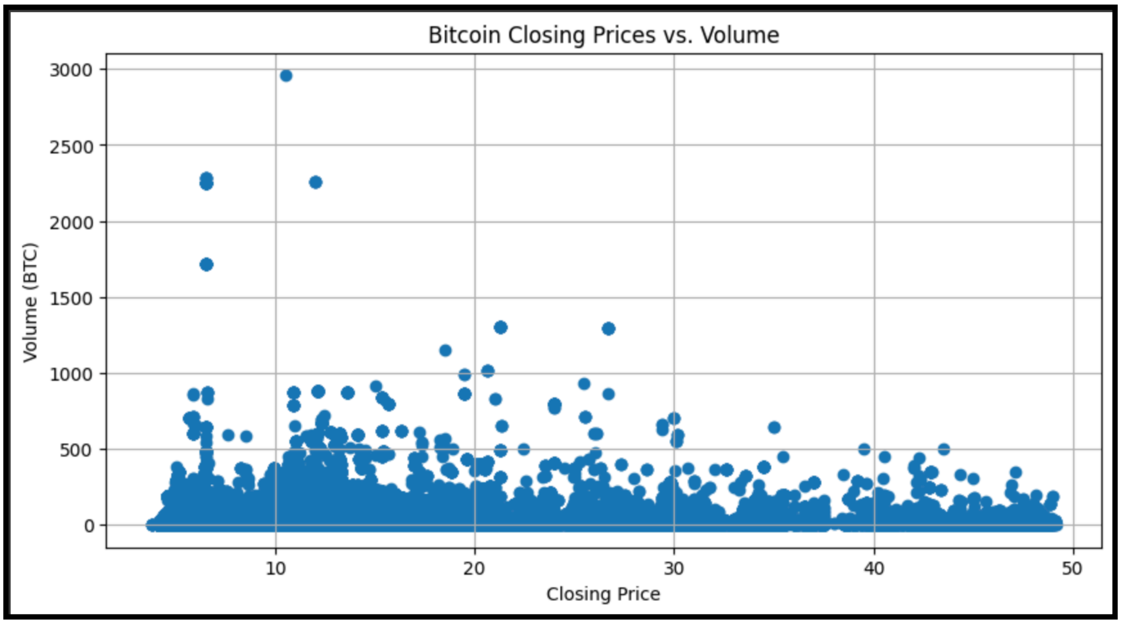
In order to facilitate a comparison of the deep learning methods to more traditional methods we built (and optimised) an ARIMA model, as they have been extensively used in price prediction problems. The ARIMA forecast was created by splitting the data into 5 periods and then predicting 30 days into the future. The data was differenced before being fit with several ARIMA models. The best fit was found by auto.arima from the R forecast package.



Comparison of ARIMA , SARIMA and Gradient Boosting Regressor Model



Data Flow Diagram for whole process



Scatter Plot of Bitcoin Closing Price and Volume

Literature Survey

This table distills key insights from various Economic & CryptoCurrency studies, encompassing authors, approaches, and techniques, serving as a comprehensive reference for predicting Bitcoin prices.

Table 1: Summary of work done in the field of Cryptocurrencies

SNo	Authors	Approaches	Techniques
1.	S. Nakamoto	Leveraging the concept of electronic cash system in early years of Bitcoin innovation.	Adoption of online payments to be sent directly from one party to another without going through a financial institution.
2.	M. Briere, K. Oosterlinck and A. Szafarz	To get return from Virtual Currency and take it as a tangible return. Use the concept of Portfolios and Diversify it along with the Bitcoins.	Use the Weekly data from 2010-2013, analyse the BTC investment from the standpoint of time series change of the price of Bitcoin.
3.	I. Kaastra and M. Boyd	Detecting financial and economic time series using a 8-step procedure to design a neural network forecasting model including tradeoffs in parameter selection.	Utilising Artificial Neural Networks as a function approximates in the field os computer science & engineering. Use Neural Network for pattern recognition.
4.	H. White	Detecting economic predictions using the concept of Neural Networks, study the case of daily ibm stock return.	Employing the concept of Neural-Network Modelling and Machine Learning techniques to search for and decode nonlinear regularities in asset price movements. Focus on IBM common stock daily returns.
5.	B. Scott	Proposing an academic database for Bitcoin and understand the important parameters required for visualising and analysing the time series data for Bitcoin Closing Price.	Experimentally analyse each parameter from the Bitcoin historical data and maps those important features on which the Bitcoin Price mostly depends such Opening price, High price, Low price and the timestamp.

SNo	Authors	Approaches	Techniques
6.	M. D. Rechenthin	Using Machine Learning classification techniques for the analysis and prediction high-frequency stock direction. Explores the predictability in the market and then design a decision support framework to provide suggestive indications of future stock price direction along with associated probability.	Utilising Machine Learning models on most up-to date price data and use the Accuracy or AUC with streaming high frequency stock data to make a plot of Bitcoin Price Changing along with different time stamps. Use Machine Learning along with drifting concept which is to assume that concept drift occurs ad builds that into the model.
7.	D. Shah & K. Zhang	Identifying the Bitcoin concept by Bayesian Regression and use its efficacy for predicting price variation of Bitcoin .	Utilising empirical data for Bayesian Regression as proxy to perform Bayesian Inference. Use Latent Source model for Binary Classification and utilise it for predicting real valued quantity, i.e. the price of Bitcoin.
8.	G. H. Chen, S. Nikolov and D. Shah	Improve time series classification using the latent source for non-parameters which includes visualising the possible patterns in many time series datasets.	Utilising a Latent Source Model Hypothesis and experimentally use Hit & trials on the synthetic data over 200 latent sources . And use the concept of Binary Classification to assign +1 and -1 labels to time series data.
9.	I. Georgoula, D. Pournarakis, C. Bilanakos, D. N. Stir	Detecting Bitcoin prices by using the time-series and sentiment analysis and calculating determinants. Also analyse the relationship between Bitcoin Prices and fundamental variables.	Techniques used are Sentiment Analysis and Support Vector Machines to derive the relationship between Bitcoin Prices and economic variables, technological factors & measurement of collective mood derived from Historical data.

10.	M. Matta, I. Lunesu and M. Marchesi	PipeLine the Bitcoin Price Data after visualising social & web-search media. Studied the existing relationship between Bitcoin's trading volume and the trading volume of Google search engine. Consider that the frequency of searches of terms about Bitcoin are significant.	Make use of search queries from the news data on Bitcoin and Cryptocurrencies to collect useful information for financial applications. Achieved significant correlation values, & demonstrate search volumes power to anticipate trading volumes of Bitcoin currency.
S . N	Authors	Approaches	Techniques
11 .	B. Gu, P. Konana, A. Liu, B. Rajagopalan & J. Ghosh	Identify information in Stock exchange message boards and its implications for stock market efficiency.	Presents an intuitive approach to identify and aggregate information in stock message boards. Weigh each post's recommendations by its authors credibility based on accuracy of his past posts. As weighted average recommendation of a stock exchange board has prediction power over future excessive returns of the stock.
12 .	R. Deflin Vidal	Explored the fractal nature of Bitcoin and collect the Evidence from wavelet spectra. Despite the asset's novelty and high volatility, evidence from the wavelet power spectra shows clear dominance of specific investment horizons during periods of high volatility.	Use the concept of continuous wavelet transform performed Bitcoin's historical returns. observe the presence of fractal dynamics in the asset's behaviour. Wavelet analysis is a method to decompose a time series into several layers of time scales, making it possible to analyse how the local variance, or wavelet power, changes both in the frequency and time domain.
13 .	L. Kristoufek	Explored the main drivers of the Bitcoin Price and use the evidence from wavelet coherence analysis. Utilise the wavelets methodology and use the Blockchain technology which freely provides very detailed series of financial bitcoin	A wavelet is a complex valued integrable function , with scale s and location u at time t. It provides the way by which any time series can be reconstructed back from its wavelet transform.
14	Y. Yoon & G. Swales	Predicting stock price performance using Neural Network approach.	Used Multivariate analytical techniques using both quantitative and qualitative variables . A Neural Network was demonstrated to address complex problems.

15.	T. Koskela, M. Lehtokanglas, J. Sarriren and K. Kaski	Implement time series prediction with multilayer perceptron fir and elm neural networks.	Use the concept of MLP to analyse one observation at a time. The output from each layer in RNN stored in context layer.

Project Outcomes

Project Outcomes: -

Accurate price Forecasts: - The primary outcome is the model's ability to generate accurate predictions of Bitcoin Prices over different time horizons, ranging from short-term fluctuations to long term trends.

Decision Support: - The model serves as a decision support tool for investors & traders, enabling them to make more informed decisions regarding buying, selling or holding Bitcoin assets. By incorporating the model's predictions into their investment strategies, stakeholders can better manage risk and capitalise on market opportunities.

To get Market Insights: - Through the analysis of historical data & model's predictions, the project generates valuable insights into Bitcoin market dynamics, including patterns, trends & factors influencing price movements. These insights contribute to a deeper understanding of the cryptocurrency market and can inform future research & investment strategies.

Evaluation of Model's performance: - The project outcomes include an evaluation of the model's performance using appropriate metrics such as accuracy, precision, recall, and mean absolute error. This assessment provides validation of the model's effectiveness in predicting Bitcoin prices and guides further refinement & optimisation efforts.

Risk Management: - The model facilitates risk management by providing insights into the uncertainty associated with price predictions. Understanding the level of uncertainty allows investors to adjust their risk exposure accordingly & implement strategies to mitigate potential losses.

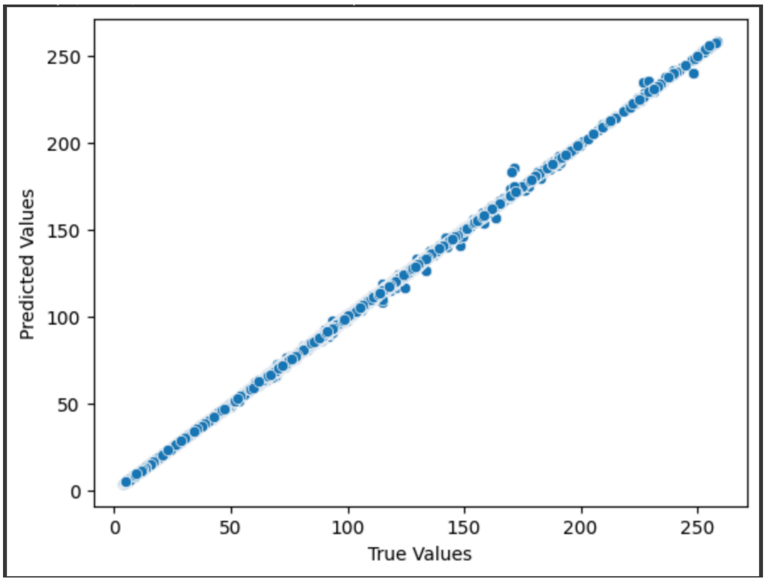
Result & Analysis

LSTM achieved the highest accuracy while the RNN achieved the lowest RMSE. The ARIMA prediction performed poorly in terms of accuracy and RMSE. Upon analysis of the ARIMA forecast, it predicted the price would gradually rise each day. There were no false positives from the model. One reason for this may be due to the class imbalance in predictive portion of the ARIMA forecast (the price tends to always increase). This contributed to the specificity and precision being so high (specificity, precision= 100%). This does not necessarily suggest good overall performance, but rather that it does a decent job at identifying price direction change(s).

Based on the results on the validation data it is apparent that all models struggled to effectively learn from the data. On training data, the model reduced error to below 1%. On validation data the LSTM achieved error of 8.07% while the RNN achieved error of 7.15%. The 50.25% and 52.78% accuracy achieved by the neural network models is a marginal improvement over the odds one has in a binary classification task (price up vs. down), i.e. 50%. The RNN was effectively of no use when using a temporal length over 50 days. In contrast, the LSTM performed better in the 50 to 100 day range with 100 days producing the best performance.

To evaluate the training of the models, Table II compares different models. The CPU utilised was a Intel Core i7 2.6GHz. The GPU used was a NVIDIA GeForce 940M 2GB. Both were running on Ubuntu 14.04 LTS installed on a SSD. For comparability the same batch size and temporal length of 50 were chosen for both the RNN and LSTM. The GPU considerably outperformed the CPU. In terms of overall training time for both networks, the GPU trained 67.7% faster than the CPU. The RNN trained 58.8% faster on the GPU while the LSTM trained 70.7% faster on the GPU. From monitoring performance in Glances, the CPU spread the algorithm out over 7 threads. The GPU has 384 CUDA cores which provide it with greater parallelism. These models were relatively quite small in terms of data with two layers. For deeper models with more layers or bigger datasets the benefits of implementing on a GPU will be even greater.

That the LSTM takes longer to train than the RNN with the same network parameters, may be due to the increased number of activation functions, and thus an increased number of equations to be performed by the LSTM. Due to the increased computation, this raises the question of the value of using an LSTM over an RNN. In financial market prediction small margins can make all the difference. As a result of this the use of an LSTM is justified. In other areas the slight improvement in terms of performance isn't justifiable for the increase in computation.



SNS plot between Predicted and Actual Values

Based on the provided RMSE values, we can draw the following conclusions:

ARIMA Model: The ARIMA model has an RMSE of 14490.55. This indicates that, on average, the predictions of the ARIMA model deviate from the actual values by approximately 14490.55 units.

SARIMA Model: The SARIMA model has an RMSE of 13692.22. This indicates that it performs slightly better than the ARIMA model, with lower prediction errors.

LSTM Model: The LSTM model has an RMSE of 19658.40. It appears to have higher prediction errors compared to the ARIMA and SARIMA models.

Gradient Boosting Regressor Model: The Gradient Boosting Regressor model has an RMSE of 14573.20. It performs similarly to the ARIMA model in terms of prediction errors.

Based on these results, the SARIMA model seems to be the best performer among the tested models, as it has the lowest RMSE value. However, it's important to note that RMSE alone may not provide a complete assessment of model performance. It's recommended to consider other evaluation metrics and also assess the models based on their suitability for the specific problem and dataset.

Additionally, it's worth mentioning that the chosen models and their hyperparameters can still be further optimized to potentially improve their performance.

Overall, the conclusion would be that the SARIMA model shows promise in predicting Bitcoin prices, but further analysis and experimentation are required to develop a more accurate and reliable forecasting model.

```
Test Set Evaluation:
_________________
MAE 0.016662316407998423
MSE 0.012003232602479114
RMSE 0.10955926525163956
R2 square 0.999993543263879
------------------------------------------
Train Set Evaluation:
___________________
MAE 0.016604222408770786
MSE 0.012312349951641665
RMSE 0.11096102897703168
R2 square 0.9999933918286871
------------------------------------------
```

Comparison of Regression metrics on train and test set

	Classifier_Name	Accuracy
1	Logistic_Regression	[0.29683634216028987]
2	Random_forest	[0.7508898785295854]
3	Decision_Tree	[0.9998244185860907]
4	LSTM	[0.0]
5	Artificial_Neural_Network	[0.0]

	Classifier_Name	Accuracy
1	Logistic_Regression	[0.29683634216028987]
2	Random_forest	[0.7508898785295854]
3	Decision_Tree	[0.9998244185860907]
4	LSTM	[0.0]
5	Artificial_Neural_Network	[0.0]
6	K_N_N_Classifier	[0.9969033823365098]

Accuracy of Different Machine Learning Classifiers on Bitcoin Dataset

Conclusion and Future Scope

Deep learning models such as the LSTM are evidently effective for Bitcoin prediction with the LSTM more capable for recognising longer-term dependencies. However, a high variance task of this nature makes it difficult to transpire this into impressive validation results. As a result it remains a difficult task. There is a fine line between overfitting a model and preventing it from learning sufficiently. Dropout is a valuable feature to assist in improving this. However, despite using Bayesian optimisation to optimise the selection of dropout it still couldn't guarantee good validation results. Despite the metrics of sensitivity, specificity and precision indicating good performance, the actual performance of the ARIMA forecast based on error was significantly worse than the neural network models. The LSTM outperformed the marginally, but not significantly. However, the LSTM takes considerably longer to train.

The performance benefits gained from the parallelisation of machine learning algorithms on a GPU are evident with a 70.7% performance improvement for training the LSTM model. Looking at the task from purely a classification perspective it may be possible to achieve better results. One limitation of the research is that the model has not been implemented in a practical or real time setting for predicting into the future as opposed to learning what has already happened. In addition, the ability to predict using streaming data should improve the model. Sliding window validation is an approach not implemented here but this may be explored as future work. One problem that will arise is that the data is inherently shrouded in noise.

In terms of the dataset, based on an analysis of the weights of the model the difficulty and hash rate variables could be considered for pruning. Deep learning models require a significant amount to data to learn effectively. If the granularity of data was changed to per minute this would provide 512, 640 data points in a year. Data of this nature is not available for the past but is currently being gathered from CoinDesk on a daily basis for future use. Finally, parallelisation of algorithms is not limited to GPU devices. Field Programmable Gate Arrays (FPGA) are an interesting alternative to GPU devices in terms of parallelisation and machine learning models have been shown to perform better on FPGA than on a GPU.

References

1. S. Nakamoto, (2008). A peer-to-peer electronic cash system. Use of distributed timestamp server on a peer-to-peer basis & use of proved work system.

2. M. Briere, K. Oosterlinck & A. Szafarz, (2013). Virtual Currency tangible return: Portfolio diversification and Bitcoins.

3. I. Kaastra & M. Boyd, (1996). Designing a neural network , Financial & Economic Time Series. Neurocomputing, vol.10, no.3, pp.215-236.

4. H. White, (1988). Economic predictions using Neural Networks. The case of daily stock returns, Neural Networks 1988. IEEE International Conference on, pp. 451-458.

5. B. Scott, (2016). Bitcoin Academic Database.

6. M. D. Rechenthin, (2014). Machine Learning Classification techniques for the analysis and prediction of high-frequency stock direction.

7. D. Shah & K. Zhang, (2014). Bayesian Regression & Bitcoin, Communication Control & Computing(Allerton). 52nd Annual Allerton Conference on, pp. 409-414.

8. G. H. Chen, S. Nikolov & D. Shah, (2013). A latent source model for non-parametric time series classification, Advances in Neural Information Processing Systems, pp. 1088-1096.

9. I. Georgoula, D. Pournarakis, C. Bilanakos, D. N. Sotiropoulos & G. M. Giggles, (2015). Using time-series & Sentiment Analysis to detect the determinants of Bitcoin Prices , at SSRN 2607167.

10. M. Matta, I. Lunesu & M. Marchesi, (2015). The predictor impact of web search media on Bitcoin Trading Volumes.

11. B. Gu, P. Konana, A. Liu, B. Rajagopalan & J.Ghosh, (2006). Identifying Information in stock message boards and its implications for stock market efficiency. Workshop on Information Systems & Economics.

12. R. Delfin Vidal, (2014). The fractal nature of bitcoin: Evidence from wavelet power spectra .

13. L. Krishtoufek, (2015). "What are the main drivers of the Bitcoin Price?", evidence from wavelet coherence analysis, PLoS, one, vol.10, no.4, pp. e0123923.

14.Y. Yoon & G. Swales, (1991). Predicting Stock Price performance: A Neural Network Approach, System Sciences. Proceedings of the Twenty-Fourth Annual Hawaii International Conference on, vol.4, pp. 156-162.

15.T. Koskela, M. Lehtokangas, J. Saarinen & K. Kaski, (1996). Time Series prediction with multilayer perceptron fir & elm neural networks, Proceedings of the World Congress on the Neural Networks, pp. 491-496