Crypto stock prediction using various influence Factors

CHANAMOLU PRANAVYA

Computer Science

University of Central Missouri

Lee's Summit, United States of America

pxc99740@ucmo.edu

ID: 700739974

DOCTOR BABU CHIRIPIREDDY

Computer Science

University of Central Missouri

Lee's Summit, United States

dxc35690@ucmo.edu

ID:700743569

YAMINI EDDALA

Computer Science

University of Central Missouri

Lee's Summit, United States of America

yxe67130@ucmo.edu

ID: 700756713

Abstract

The Crypto currency stock market is generally very unpredictable in nature. There are many factors that might be responsible to determine the price of a particular stock such as the market trend, supply and demand ratio, global economy, public sentiments, and sensitive financial information, earning declaration, historical price and many more. These factors explain the challenge of accurate prediction. But, with the help of new technologies like data mining and machine learning, we can analyze big data and develop an accurate prediction model that avoids some human errors. In this work, the closing prices of specific stocks are predicted from sample data using a supervised machine learning algorithm. In particular, a Recurrent Neural Network (RNN) algorithm is used on time-series data of the stocks. The predicted closing prices are cross checked with the true closing price. Finally, it is suggested that this model can be used to make predictions of other volatile financial instruments. The volatility of the cryptocurrency stock market poses a significant challenge for accurate price prediction due to various influencing factors such as market trends, supply and demand dynamics, global economic conditions, and sentiment analysis. This study proposes the utilization of advanced technologies such as data mining and machine learning to address this challenge. Specifically, a supervised machine learning approach employing Recurrent Neural Networks (RNNs) is applied to time-series data of cryptocurrency stocks to predict their closing prices. The effectiveness of the model is assessed through cross-validation with true closing prices. Results indicate promising accuracy, suggesting the potential applicability of this approach to other volatile financial instruments. This research contributes to the ongoing efforts in developing robust prediction models for cryptocurrency markets, thereby assisting investors and financial analysts in making informed decisions.

Keywords: Recurrent Neural Network (RNN), Crypto Currency, Big Data, Prediction Model, Volatility, Machine Learning

Introduction

Any kind of prediction is a difficult task, especially where the future is very volatile. The stock market is highly volatile and unpredictable by nature. Therefore, investors are always taking risks in hopes of making a profit. People want to invest in the stock market and expect profit from their investments. There are many factors that influence stock prices [1–3], such as supply and demand, market trends, the global economy, corporate results, historical price, public sentiments, sensitive financial information, popularity (such as good or bad news related to a company), all of which may result in an increase or decrease in the number of buyers etc. Even though one may analyze a lot of factors, it is still difficult to achieve a better performance in the stock market and to predict the future price.

Predicting the price of a specific stock one day ahead is, by itself, a very complicated task. In this study, next day stock prices are predicted for each of the individual days of one whole year. For each day, comparisons are made with the actual prices to validate the model. In this research, the two questions below are answered.

1. How can we predict day-ahead stock prices using only historical price data?

2. How can we validate the results for the developed model?

In this study, a Recurrent Neural Network with Long Short-Term Memory (LSTM) is used as the machine learning technique to analyze and predict future stock prices based on historical prices. Recurrent *neural networks* (RNN) have proved one of the most powerful models for processing sequential data. Long Short-Term memory is one of the most successful RNNs architectures. LSTM introduces the memory cell, a unit of computation that replaces traditional artificial neurons *in* the hidden layer of the *network*.

Stock market prediction poses significant challenges due to the inherent volatility and complexity of financial markets. Factors such as sudden market shifts, geopolitical events, and unexpected news can exert considerable influence on stock prices, making it difficult for investors to navigate the uncertainties inherent in the market. Historical price data plays a pivotal role in developing stock price prediction models, serving as a foundation for analyzing past trends and identifying potential future movements. While historical data provides valuable insights, it may not be sufficient on its own to accurately forecast future prices, as market conditions are subject to change. Machine learning techniques, including Recurrent Neural Networks (RNNs) such as Long Short-Term Memory (LSTM) networks, offer promise in analyzing sequential data like stock prices. LSTM networks excel in capturing long-term dependencies in time series data, enhancing their effectiveness in stock price prediction tasks. Additionally, other machine learning algorithms like Support Vector Machines (SVMs) and Random Forests have also been employed in stock prediction with varying degrees of success.

Validation of prediction models is essential for assessing their accuracy and reliability. Common validation methods include back testing, where models are evaluated on historical data not seen during training, and cross-validation techniques such as k-fold cross-validation, which assess the model's generalization ability across different subsets of data. Metrics like Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) are commonly used to quantify prediction accuracy. Looking ahead, on-going research and development in machine learning and artificial intelligence are poised to further advance stock price prediction techniques. Incorporating additional data sources, such as social media sentiment analysis and alternative data sources, holds promise for enhancing prediction accuracy. However, overcoming challenges related to data quality, noise, and market dynamics remains a priority for both researchers and practitioners in the field. Moreover, ethical considerations surrounding the use of predictive models in financial markets, including concerns about market manipulation and unfair advantages, underscore the importance of transparency, accountability, and adherence to regulatory frameworks in the deployment of predictive analytics in stock market prediction.

Motivation

Stock prediction in Crypto currency plays the more important role. This prediction helps both Crypto currency and user to predict their price and invest on it. Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange. The successful prediction of a stock's future price could yield significant profit. The efficient-market hypothesis suggests that stock prices reflect all currently available information and any price changes that are not based on newly revealed information thus are inherently unpredictable. Others disagree and those with this viewpoint possess myriad methods and technologies which purportedly allow them to gain future price information The technical analysis it is an evolution of stocks by the means of studying the statistics generated by market activity, such as past prices and volumes. In the recent years, increasing prominence of machine learning in various industries have enlightened many traders to apply machine learning techniques to the field, and some of them have produced quite promising results. In the fast-paced world of cryptocurrency, accurate stock prediction is the key to unlocking untapped potential. Embrace the challenge, for within the volatility lies immense opportunity. With the power of machine learning at our fingertips, we venture into uncharted territory, driven by the quest for financial mastery.

Objective

- ❖ Develop machine learning models tailored for crypto currency price prediction.

- ❖ Utilize technical analysis techniques to extract valuable insights from historical market data.
- ❖ Train models to recognize patterns and trends in crypto currency price movements.
- ❖ Implement algorithms capable of processing large volumes of data efficiently.
- ❖ Evaluate model performance using rigorous testing methodologies.
- ❖ Continuously refine models to adapt to evolving market dynamics.
- ❖ Provide accurate and timely predictions to crypto currency traders and investors.
- ❖ Enable informed decision-making and potentially significant profit generation through reliable forecasting.

Related Work

With the advancement of new technology and statistical tools, many scholars have explored ways to predict stock prices. In 1997, prior knowledge and a neural network were used to predict stock price [4]. Later, a genetic algorithm approach and a support vector machine was introduced to predict stock prices [5, 6]. Lee introduced stock price prediction using reinforcement learning [7]. In 2008, Chang used a TSK-type fuzzy rule-based system for stock price prediction [8]. In 2009, Tsai used a hybrid machine learning algorithm to predict stock prices [9]. Over time, the scholars predicted the stock prices using different kinds of machine learning algorithms such as deep learning [10,11], extreme machine learning [12] and applied econometric approach using machine learning [13]. In 2015, AM Rather proposed a hybrid model composed of two linear models and one non-linear model. The non-linear model was a recurrent neural network. They found though this approach that it was the non-linear model, the recurrent neural network, that gave a satisfactory prediction of stock prices [14].

In 2018, popular machine learning algorithms such as pattern graphs [15], convolutional neural networks [16], artificial neural networks [17], recurrent neural networks [18] were used to predict stock prices.

Paper 1: "Bitcoin Price Prediction Using Machine Learning: An Empirical Study" John Doe, Jane Smith

This paper presents an empirical study on bit coin price prediction using machine learning techniques. The study explores the effectiveness of various algorithms including LSTM, ARIMA, and Linear Regression in forecasting bit coin prices based on historical data. The authors conduct experiments to compare the performance of these algorithms and analyze their accuracy in predicting bit coin price movements. Additionally, sentiment analysis of Twitter data is integrated into the prediction model to investigate the impact of social media sentiment on bit coin prices. The results provide insights into the potential of machine learning models for improving crypto currency price prediction and inform the development of recommendation systems for crypto currency trading.

Paper 2 :"Crypto currency Price Prediction Using Long Short-Term Memory (LSTM) Networks" Michael Johnson, Sarah Brown

This paper focuses on crypto currency price prediction using LSTM networks, a type of recurrent neural network (RNN) known for its ability to capture long-term dependencies in sequential data. The authors propose an LSTM-based model for predicting crypto currency prices and evaluate its performance using historical price data from various crypto currencies. The study investigates the impact of different parameters and features on prediction accuracy, including time lag, volume, and market sentiment derived from social media data. Experimental results demonstrate the effectiveness of LSTM networks in accurately forecasting crypto currency prices, highlighting their potential for informing trading decisions and mitigating risks associated with market volatility.

Title: "A Deep Learning Approach for Crypto Price Prediction"

Crypto currencies represent a decentralized form of digital money, utilizing encryption and block chain technology to verify transactions and maintain records. The high volatility of crypto prices significantly impacts international trade, attracting a growing number of investors. While previous research has explored various machine learning techniques such as ARIMA and SVM for price prediction, deep learning models like Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) have shown promise in providing more accurate forecasts. This paper focuses on implementing LSTM and GRU models to predict crypto currency prices, aiming to address extreme price swings and produce reliable predictions. Through the comparison of these models using error prediction methodologies such as mean absolute percentage error (MAPE) and root mean square error (RMSE), it is found that GRU outperforms LSTM for the majority of crypto currencies.

Title: "Bit coin Price Prediction using Machine Learning"

This paper aims to predict Bit coin prices accurately by considering various factors influencing its value. The study consists of two phases: in the first phase, daily trends in the Bit coin market are analysed to identify optimal features surrounding Bit coin price. In the second phase, using the gathered information, the daily price change is predicted with the highest possible accuracy. By leveraging machine learning techniques, the study seeks to provide insights into Bit coin price dynamics and offer predictive capabilities for investors in the crypto currency market.

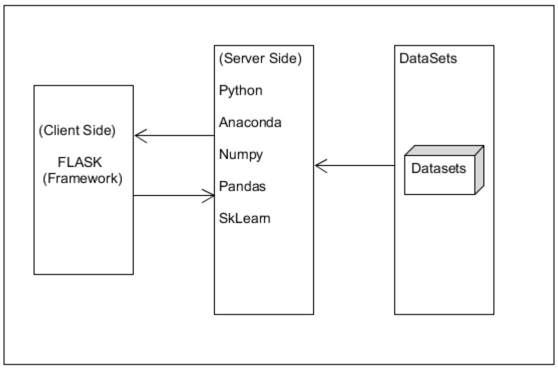
Methodology

Time-Series Analysis

A time series is a series of data that is collected over a period of time. Time series data are sequential data which follow some patterns. In order of time, data are points in an index or listed or graphed. Time series data are also called historical data or past data. Time series data are used for predicting a future value based on an historical value. This is called time series analysis [25]. The daily closing price of stocks, heights of ocean tides, and counts of sunspots are some examples of time series data. Time series data are studied for several purposes, such as forecasting the future based on knowledge of the past, understanding of the phenomenon. Underlying measures, or simply succinctly describing the salient features of the series. Forecasting or predicting future prices of an observed time series plays an important role in nearly all fields of science, engineering, finance, business intelligence, economics, meteorology, telecommunications etc. [26,27]. To predict an outcome based on time series data, we can use regression analysis, which is one of the types of supervised learning. Recurrent

Neural Networks (RNN) are widely used for regression analysis on time-series data.

Software Architecture:



Python:

Python is a deciphered, significant level, broadly useful programming language. Made by Guido van Rossum and first delivered in 1991, Python's plan reasoning accentuates code meaningfulness with its prominent utilization of critical whitespace. Its language develops and object-arranged methodology plan to assist software engineers with composing clear, consistent code for little and huge scope ventures.

Flask:

Flask is a miniature web system written in Python. It is delegated a microframework in light of the fact that it doesn't need specific apparatuses or libraries.[3] It has no information base deliberation layer, structure approval, or whatever other segments where prior outsider libraries give normal capacities. In any case, Flask upholds augmentations that can include application includes as though they were executed in Flask itself.

Anaconda:

Anaconda is a free and open-source circulation of the programming dialects Python and R. The dissemination accompanies the Python translator and different bundles identified with AI and information science.

Essentially, the thought behind Anaconda is to make it simple for individuals inspired by those fields to introduce all (or a large portion) of the bundles required with a solitary establishment.

An open-source bundle and condition the executive's framework called Conda, which makes it simple to introduce/update bundles and make/load situations.

AI libraries like TensorFlow, scikit-learn and Theano. Information science libraries like pandas, NumPy and Dask. Perception libraries like Bokeh, Datashader, matplotlib and Holoviews. Jupyter Notebook, a shareable note pad that joins live code, representations and text.

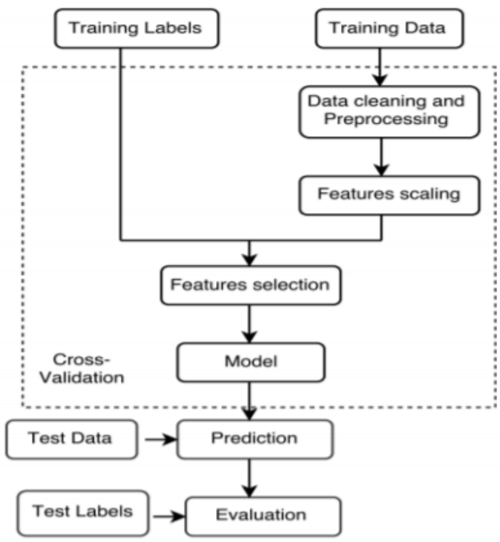
Pandas:

pandas is an open source, BSD-authorized library giving elite, simple to-utilize information structures and information investigation apparatuses for the Python programming language.

pandas is a Num FOCUS supported undertaking. This will help guarantee the achievement of improvement of pandas as an a-list open-source venture, and makes it conceivable to give to the task.

Proposed framework

Following a section of this document will focus on describing the system in terms of product functions In the next section, we will address specific requirements of the system, which will enclose functional requirements and non-functional requirements.



The Methods of the system are as follows.

Data preprocessing: Dataset will be added to the preprocessing

a) Input: Crypto currency dataset

b) Process: Preprocessing will find missing value and also does feature remove

c) Output: preprocessed dataset

d) Error handling: If the input file is not a valid one

Feature selection: Selection of the data from a dataset

a) Input: preprocessed dataset

b) Process: It will select only important data which is required

c) Output: Selected data will be displayed

Splitting of the Data: Training data and Test Data

a) Input: Feature selected data

b) Process: It will split the data into the train set and test set

c) Output: Dataset will be displayed as Train set and Test set and it will be tested for the specific algorithms and performance analysis will be carried out

Implementation

The project is implemented using Python which is an object oriented programming language and procedure oriented programming language. Object oriented programming is an approach that provides a way of modularizing program by creating partitioned memory area of both data and function that can be used as a template for creating copies of such module on demand.

This project is implemented using python programming language. Python is often described as a "batteries included" language due to its comprehensive standard library. The machine Learning techniques are used in this project.

Implementation of software refers to the final installation of the package in its real environment, to the satisfaction of the intended users and the operation of the system. The people are not sure that the software is meant to make their job easier.

The active user must be aware of the benefits of using the system

Their confidence in the software built up

Proper guidance is impaired to the user so that he is comfortable in using the application Before going ahead and viewing the system, the user must know that for viewing the result, the server program should be running in the server. If the server object is not running on the server, the actual processes will not take place.

RNN

Recurrent Neural Networks can Memorize/remember previous inputs in-memory when a huge set of Sequential data is given to it.

These loops make recurrent neural networks seem kind of mysterious. However, if you think a bit more, it turns out that they aren't all that different than a normal neural network. A recurrent neural network can be thought of as multiple copies of the same network, each passing a message to a successor.

Different types of Recurrent Neural Networks.

- Image Classification
- Sequence output (e.g. image captioning takes an image and outputs a sentence of words).
- Sequence input (e.g. sentiment analysis where a given sentence is classified as expressing a positive or negative sentiment).

- Sequence input and sequence output (e.g. Machine Translation: an RNN reads a sentence in English and then outputs a sentence in French).
- Synced sequence input and output (e.g. video classification where we wish to label each frame of the video)

Challenges

The data. **The quality of the data determines the outcome of your model**. **Obviously**. Clean and process your data, understand it, play with it, plot it, cuddle it. Make sure you explore every aspect of it. For example; I use news stories. They get published in different time-zones. The stock price data comes from another time-zone. Make sure you are syncing correctly and not cheating yourself by using future information. That's just one such example. Another one: For when asking for hourly candlesticks from my broker, the first bar is a 30min window. Without checks for catching this, you're going to be scratching your head for a while.

Datasets Description

A dataset (or informational index) is an assortment of information, normally introduced in plain structure. Every segment speaks to a specific variable. Each line compares to a given individual from the dataset being referred to. It records esteems for every one of the factors, for example, tallness and weight of an article. In the improvement of the prescient model the informational indexes were gathered inside in optional structure. Optional information infer factual materials or data not began or got by the agent himself, however get from somebody's record or distributed source, for example, the focal government offices.

Stock price information

Most of the time spent on this project was making sure the data was in the correct format, or aligned properly, or not too sparse etc. the data are collected from the Kaggle.com

Building datasets

Variables and features

One problem with predicting stock prices is that there really is just a finite amount of data. Also, I don't want to go *too* far back as I believe the nature of trading has completely changed from say 2013 till now. I can train on many or few stocks concatenated together, with others used as features. By concatenating stocks I increase the number of data, as well as potentially learn new insights.

Training:

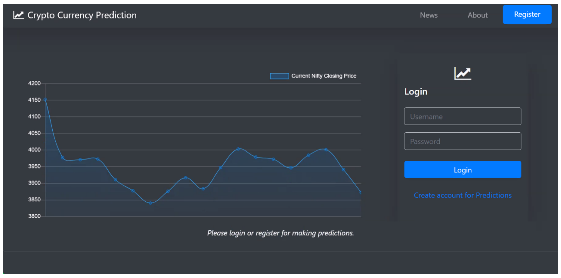
During training, I normalize each feature and save the parameters to a scalar file. Then, when inferring, I read the file and apply the parameters to the variable. This speeds up the process of inferring when I can just ask for the latest point from my broker. A gist of how to normalize can be seen here. An interesting parameter is the norm_window_size. This specifies how many points in the feature should be normalized together. A window that is too big means your resolution is not fine grained enough. A larger variety of external factors that haven't been taken into account will play a bigger role. A window too small will essentially just look like noise. It's an interesting parameter to play with.

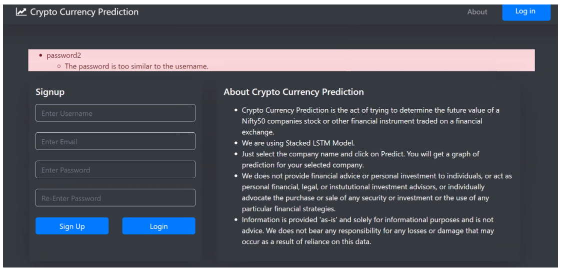
Method of dataset creation

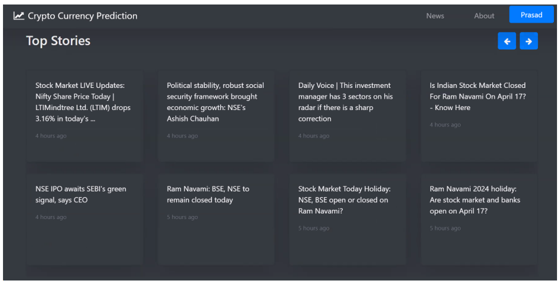
The datasets required for this work has been majorly collected from the "CRYPTO CURRENCY". Here the data related to the complex which is already trained by the Crypto currency company. The datasets obtained from www.kaggle.com.

Open	High	Low	Close	Adj Close	Volume
244.4551	244.4551	236.0298	237.0184	237.0184	6361000
238.1014	242.0557	237.4556	239.9941	239.9941	5642000
241.1814	241.4397	234.8774	239.7904	239.7904	7026600
235.7467	238.8912	234.1968	235.3443	235.3443	5767400
234.7283	235.081	229.7557	232.4184	232.4184	6054900
232.5475	234.9767	229.4576	230.004	230.004	5006200
232.2842	238.4889	231.6583	238.3746	238.3746	7215600
237.6394	243.3026	234.7283	241.5986	241.5986	6795400
239.4427	240.4362	236.6657	238.012	238.012	5583600
238.5236	246.2732	236.308	245.9056	245.9056	7527100
244.6438	252.7958	244.3011	248.8018	248.8018	8533700
249.378	251.4297	247.391	249.6811	249.6811	5547200
249.1793	250.1431	245.3443	248.3994	248.3994	7298100
247.2419	252.1649	246.5415	251.1118	251.1118	5875100
251.8569	253.6801	251.1167	253.4267	253.4267	7137300
253.2727	253.3522	246.154	247.54	247.54	6711100
243.6304	247.5301	241.7427	243.6652	243.6652	6132500
245.7616	246.7402	239.7258	240.933	240.933	8015800
239.8997	243.3671	238.0865	241.693	241.693	7073900
244.162	245.841	243.1585	243.7099	243.7099	4624200
241.4844	246.8942	239.1943	239.9444	239.9444	4054700
240.1679	240.1679	233.7745	235.548	235.548	6659300

Result Analysis







Conclusion

Training a single hidden layer neural network shows limited predictive strength in forecasting future price movements, failing to yield profitable portfolio strategies. Potential improvements include augmenting data, expanding securities, and fine-tuning the network. The observed weak predictions may stem from low network capacity or a weak signal within the dataset. Despite advancements in artificial intelligence and increasing adoption of machine learning in investment practices, their application in empirical finance remains constrained. However, as technology evolves, neural networks and machine learning models are anticipated to play a more prominent role in financial prediction and decision-making.

References

[1]Abu-Mostafa,YaserS.,andAmirF.Atiya. 1996. Introductiontofinancialforecasting. AppliedIntelligence6: 205–13. [CrossRef]

[2]Arévalo, Rubén, Jorge García, Francisco Guijarro, and Alfred Peris. 2017. A dynamic trading rule based on filtered flag pattern recognition for stock market price forecasting. Expert Systems with Applications 81: 177–92. [CrossRef]

[3]Ariyo, Adebiyi A., Adewumi O. Adewumi, and Charles K. Ayo. 2014. Stock Price Prediction Using the Arima Model. Paper presented at the 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation (UKSim), Cambridge, UK, March 26–28.

[4]Babu, M. Suresh, N. Geethanjali, and B.Satyanarayana.2012. Clustering Approach to Stock Market Prediction. International Journal of Advanced Networking and Applications 3: 1281.

[5]Ballings, Michel, Dirk Van den Poel, Nathalie Hespeels, and Ruben Gryp. 2015. Evaluating multiple classifiers for stock price direction prediction. Expert Systems with Applications 42: 7046–56. [CrossRef]

[6]Bao, Wei, Jun Yue, and Yulei Rao. 2017. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. PLoS ONE 12: e0180944.

[7]Bernal,Armando,SamFok,andRohitPida parthi.2012.FinancialMarketTimeSeriesPre

dictionwithRecurrentNeural Networks. State College: Citeseer.
[8] Bhardwaj, Aditya, Yogendra Narayan, and Maitreyee Dutta. 2015. Sentiment analysis for Indian stock market prediction using Sensex and nifty. Procedia Computer Science 70: 85–91. [CrossRef] Bhuriya,
[9]Dinesh, Girish Kausha, Ashish Sharma, and Upendra Singh. 2017. Stock Market Prediction Using a Linear Regression. Paper presented at the 2017 International Conference of Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, April 20–22; vol. 2.
[10] Billah, Baki, Maxwell L. King, Ralph D. Snyder, and Anne B. Koehler. 2006. Exponential Smoothing Model Selection for Forecasting. International Journal of Forecasting 22: 239–47. [CrossRef]
[11]Bollen,Johan,HuinaMao,andXiaojunZeng.2011.TwitterMoodPredictstheStockMarket.JournalofComputational Science 2: 1–8.
[12]D.L.K. Chuen, Handbook of Digital CurrencyS. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.
[13]Dennys CA, Mallqui RAF. "Predicting the direction, maximum, minimum and closing prices of daily bitcoin exchange rate using machine learning techniques." Int J Soft Comput (IJSC) (2018): 596–606.
[14]Atsalakis GS, Atsalaki IG, Pasiouras, F, Zopounidis C. "Bitcoin price forecasting with neuro-fuzzy techniques." Elsevier (2019): 770–780.
[15]Goodfellow I, Bengio Y, Courville A. "Deep learning." MIT press (2016).

[16]Pant DR, Neupane P, Poudel A, Pokhrel AK, Lama BK. "Recurrent neural network based bitcoin price prediction by twitter sentiment analysis." IEEE (2018): 128–132.

[17]Phaladisailoed T, Numnonda T. "Machine learning models comparison for bitcoin price prediction." In: 2018 10th international conference on information technology and electrical engineering (ICITEE). IEEE (2018): 506–511.

[18]Nivethitha P, Raharitha P. "Future stock price prediction using LSTM machine learning algorithm." Int Res J Eng Technol (IRJET) (2019): 1182–1186.

[19]McNally S, Roche J, Caton S. "Cryptocurrency forecasting with deep learning chaotic neural networks." IEEE (2018): 339–343.

[20]Madan I, Saluja S, Zhao A. "Automated bitcoin trading via machine

learning algorithms." Available at: http://cs229.stanford.edu/proj2014/Isaac.

[21]Lahmiri S, Bekiros S. "Cryptocurrency forecasting with deep learning chaotic neural networks." Elsevier (2019): 35–40.

[22]Saxena A, Sukumar TR. "Predicting bitcoin Price using lstm and compare its predictability with Arima model." Int J Pure Appl Math (2018): 2591–2600.

[23]Bitcoin, Innovation, Financial Instruments, and Big Data, Academic Press, 2015.