MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR

(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)

Final Year Internship Report

on

Machine Learning Intern at Praedico Global Research Pvt.Ltd.

Submitted By:

Ram Shararn Goyal

0901CS181079

Faculty Mentor:

Mr. Arun Kumar

Assistant Professor, Dept. Of CSE

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE
GWALIOR - 474005 (MP) est. 1957

MAY-JUNE 2022

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR

(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)

Machine Learning Intern at Praedico Global Research Pvt.Ltd.

A final year internship report submitted in partial fulfillment of the requirement for the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

Submitted by:

Ram Shararn Goyal

0901CS181079

Industry Mentor:

Priyank Gupta, CTO

Submitted to:

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE

GWALIOR - 474005 (MP) est. 1957

MAY-JUNE 2022

Address: First Floor, Garima Arcade, Shinde ki Chhawani, Gwalior
Email: mail@praedicoglobalresearch.com
Website: www.praedicoglobalreserach.com

Ref.: PGR-2022/I-508

Date: 11th – May - 2022

CERTIFICATE OF INTERNSHIP

This certificate is awarded to

Mr./Miss. RAM SHARARN GOYAL

In appreciation for your accomplishments in the company as an intern

(Position titled-***"Neural Network Developer"***)

at **Praedico Global Research Pvt. Ltd.,**

from **Jan 10th, 2022 to May 10th, 2022.**

We take this opportunity to wish you a long, happy and successful career.

Authorized Signatory
Praedico Global Research Pvt. Ltd.

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR

(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)

CERTIFICATE

This is certified that **Ram Sharan Goyal** (0901CS181079) has submitted the Internship report titled **Machine Learning Intern at Praedico Global Research Pvt. Ltd** of the work he has done under the mentorship of Mr.Arun Kumar>, in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering from Madhav Institute of Technology and Science, Gwalior.

Mr.Arun Kumar
Faculty Mentor
Assistant Professor
Computer Science and Engineering

Dr. Manish Dixit
Professor and Head,
Computer Science and Engineering

Dr. Manish Dixit
Professor & HOD
Department of CSE
M.I.T.S. Gwalior

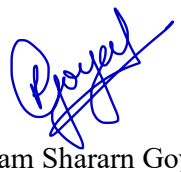
MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR

(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)

DECLARATION

I hereby declare that the work being presented in this Internship report, for the partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in CSE at Madhav Institute of Technology & Science, Gwalior is an authenticated and original record of my work under the mentorship of **Mr. Arun Kumar**, **Assistant Professor**, Department of CSE.

I declare that I have not submitted the matter embodied in this report for the award of any degree or diploma anywhere else.



Ram Shararn Goyal
0901CS181079
IV Year,
Computer Science and Engineering

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR

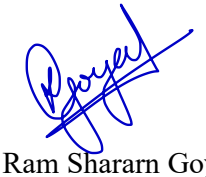
(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)

ACKNOWLEDGEMENT

The full semester internship has proved to be pivotal to my career. I am thankful to my institute, **Madhav Institute of Technology and Science** to allow me to continue my disciplinary/interdisciplinary internship as a curriculum requirement, under the provisions of the Flexible Curriculum Scheme (based on the AICTE Model Curriculum 2018), approved by the Academic Council of the institute. I extend my gratitude to the Director of the institute, **Dr. R. K. Pandit** and Dean Academics, **Dr. Manjaree Pandit** for this.

I would sincerely like to thank my department, **Department of Computer Science and Engineering, for allowing** me to explore this internship. I humbly thank **Dr. Manish Dixit**, Professor and Head, Department of Computer Science and Engineering, for his continued support during the course of this engagement, which eased the process and formalities involved.

I am sincerely thankful to my faculty mentors. I am grateful to the guidance of **Mr.Arun Kumar Assistant Professor,** Department of Computer Science and Engineering, for his continued support and close mentoring throughout the internship. I am also very thankful to the faculty and staff of the department.



Ram Shararn Goyal
0901CS181079
IV Year,
Computer Science and Engineering

ABSTRACT

I carried out my internship at Praedico Global Research Pvt Ltd. It offers internship opportunities to the students in need of education about software technology. The purpose of the program is to fulfill the core equipment for the award of a Bachelor's Degree in Technology to get a practical aspect of the of the theoretical work studied at the university and to understand the operations in the IT sector and to enable students to gain experience in different tasks. In this Internship, we learn various aspects of Machine Learning and we attempt to implement a machine learning approach to predict cryptocurrency prices.I have learnt a lot about basic machine learning algorithms and also about data visualization, data analysis and data optimization and it's various techniques which we use in machine learning and how important it is to make a machine learning model successful . Machine learning is effectively implemented in forecasting cryptocurrency prices. The objective is to predict cryptocurrency prices in order to make more informed and accurate investment decisions. We propose a cryptocurrency price prediction system that integrates mathematical functions, machine learning, and other external factors for the purpose of achieving better cryptocurrency prediction accuracy and issuing profitable trades.

सारः

मैंने प्रेडिको ग्लोबल रिसर्च प्राइवेट लिमिटेड में अपनी इंटर्नशिप पूरी की। यह सॉफ्टवेयर प्रौद्योगिकी के बारे में शिक्षा की आवश्यकता वाले छात्रों को इंटर्नशिप के अवसर प्रदान करता है। कार्यक्रम का उद्देश्य विश्वविद्यालय में अध्ययन किए गए सैद्धांतिक कार्य के व्यावहारिक पहलू को प्राप्त करने और आईटी क्षेत्र में संचालन को समझने और छात्रों को लाभ प्राप्त करने में सक्षम बनाने के लिए प्रौद्योगिकी में स्नातक की डिग्री के पुरस्कार के लिए मुख्य उपकरण को पूरा करना है। विभिन्न कार्यों में अनुभव। इस इंटर्नशिप में, हम मशीन लर्निंग के विभिन्न पहलुओं को सीखते हैं और हम क्रिप्टोकुरेंसी कीमतों की भविष्यवाणी करने के लिए मशीन लर्निंग दृष्टिकोण को लागू करने का प्रयास करते हैं। मैंने बुनियादी मशीन लर्निंग एल्गोरिदम और डेटा विज़ुअलाइज़ेशन, डेटा विश्लेषण और डेटा ऑप्टिमाइज़ेशन और इसकी विभिन्न तकनीकों के बारे में बहुत कुछ सीखा है। जिसका उपयोग हम मशीन लर्निंग में करते हैं और मशीन लर्निंग मॉडल को सफल बनाना कितना महत्वपूर्ण है। क्रिप्टोक्यूरेंसी कीमतों की भविष्यवाणी में मशीन लर्निंग को प्रभावी ढंग से लागू किया गया है। इसका उद्देश्य अधिक सूचित और सटीक निवेश निर्णय लेने के लिए क्रिप्टोक्यूरेंसी कीमतों की भविष्यवाणी करना है। हम एक क्रिप्टोक्यूरेंसी मूल्य भविष्यवाणी प्रणाली का प्रस्ताव करते हैं जो बेहतर क्रिप्टोक्यूरेंसी भविष्यवाणी सटीकता प्राप्त करने और लाभदायक ट्रेड जारी करने के उद्देश्य से गणितीय कार्यों, मशीन सीखने और अन्य बाहरी कारकों को एकीकृत करती है।

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
LSTM	Long short term memory
RNN	Recurrent neural network
CSV	Comma separated value
Btc	Bitcoin
Eth	Ethereum
MAPE	Mean absolute Percentage error
LTC	Litecoin
LGBM	Light Gradient Boosting Method
ANN	Artificial Neural Network

Chapter 1: INTRODUCTION

The word "cryptocurrency" refers to a sort of digital currency utilized in financial transactions. It is encrypted, which prevents counterfeiting and double-spending. The ability of virtual money to be traded via cryptographic methods separates it from conventional currencies issued by central authorities or central banks. The second aspect is that it is built on blockchain technology, which is incredibly complicated and aims to store data in a manner that makes it difficult or impossible to change, hack, or cheat the system. Bitcoin has started to carve out a niche for itself, which may assist or impede the broader adoption of cryptocurrencies. Cryptocurrencies are still in their infancy, and it is hard to predict whether they will ever be widely accepted in global markets. Bitcoin, the most well-known cryptocurrency, was founded in 2009 and was the only Blockchain-based cryptocurrency for more than two years. However, there are over 5000 currencies and 5.8 million active users on the bitcoin market. Bitcoin has lately attracted a lot of interest in the fields of economics, cryptography, and computer science due to its core nature of combining encryption technology with monetary units.

1.1 MOTIVATION FOR WORK

The advent of cryptocurrency as a new sort of asset that has emerged as a result of developments in financial technology has created a substantial potential for research. Cryptocurrency is a decentralized digital money that uses cryptography for security. Cryptocurrency projections are difficult because of price volatility and dynamism. Hundreds of cryptocurrencies are in use today all around the world. This research presents three distinct recurrent neural network (RNN) algorithms for predicting the prices of three major cryptocurrencies. These cryptocurrencies are Bitcoin (BTC), Litecoin (LTC), and Ethereum (ETH) (ETH). The models produce excellent forecasts based on the mean absolute percentage error (MAPE). According to the findings, the Light gradient boosting method (LGBM) outperformed the long short-term memory (LSTM) and bidirectional LSTM (bi-LSTM) models in all aspects of bitcoin prediction. As a result, it is the best algorithm for the job. These models are valuable because they can assist investors and traders in identifying bitcoin transactions and purchases, which can have significant economic ramifications. Other elements that may influence cryptocurrency market values, such as social networks, twitter tweets, and market cap, should be investigated in future research. Because of the fundamental nature of Bitcoin, which is to combine encryption technology with monetary units, it has recently garnered a lot of attention in the fields of economics, cryptography, and computer science. This is owing to the fact that Bitcoin is the first cryptocurrency to do so.

1.2 PROBLEM STATEMENT

Time series forecasting and modelling are critical in data analysis. Time series analysis is a type of statistics that is commonly used in econometrics and operations research. Time series is extensively used in analytics and data science. The value of crypto currencies is governed by a multitude of factors. The study's main purpose is to employ LSTM and LGBM to predict crypto currency prices.

CHAPTER 2: LITERATURE SURVEY

2.1 INTRODUCTION

Machine learning (ML) is a type of artificial intelligence that forecasts the future using historical data. Prior research has shown that ML-based models not only produce results that are almost or exactly the same as the actual result, but also improve the accuracy of the result. Machine learning includes neural networks, support vector machines, and deep learning. demonstrates that integrating cryptocurrency in a portfolio boosts efficiency in two ways. The first is to reduce standard deviation, and the second is to provide investors more options for allocation. The appropriate cryptocurrency allocation was suggested to be between 5% and 20% depending on the investor's risk tolerance. The authors focus on time series data forecasting and employ two machine learning algorithms: random forests (RF) and stochastic gradient boosting machine (SGBM) (SGBM). According to the findings, the ML ensemble technique can be used to predict Bitcoin values.

2.2 EARLIER METHODS OF FORCASTING

To limit the risks associated with the investment process, the decision-making process must make the optimal option at the right time. Previously, a hybrid cryptocurrency prediction system based on LSTM and GRU was shown, including examples of Litecoin and Monero. The data scientist used minute-sampled Bitcoin returns over three-hour periods to aggregate RV data. The heterogeneous auto-regressive realised volatility (HARRV) model with optimum lag parameters was compared to a range of machine learning algorithms for predicting future values based on previous samples, including ANN (MLP, GRU, and LSTM), SVM, and ridge regression. The findings show that the proposed method effectively predictions prices, meaning that the technique might be used to a variety of cryptocurrencies. The authors employ traditional support vector machine and linear regression techniques to forecast Bitcoin values. This research looks at a daily data time series forecast.

CHAPTER 3: TECHNOLOGIES USED IN THIS PROJECT

Data analytics has grown into a massive field with a wide range of technological ramifications. However, we will limit ourselves to technology utilised in the construction of our project.We would use Exploratory Data Analysis to analyse the dataset provided for crypto forecasting. (EDA)

3.1 EXPLORATORY DATA ANALYSIS

An exploratory analysis' principal goal is to investigate, as the name implies. The data's link to the elements that came before it is still uncertain. After the data have been analysed, exploratory analysis enables us to recognise connections, think of new hypotheses, and devise methods for resolving particular problems. As a result, it refers to the practise of conducting preliminary studies on data using summary statistics and graphical representations in order to find trends, detect anomalies, and test assumptions. In addition to this, we would utilise these many technological means to attain our goal.

3.2 DATA VISUALISATION

In order to better convey the meaning and context of data, data visualisation uses graphics. Organizations may examine trends and anomalies in real time by drilling down into the data using interactive data visualisation.

Because data is better understood when presented visually rather than mathematically, data visualisation is frequently the initial step in understanding and presenting analytics. When data is represented graphically, it is simpler to spot developing patterns, which in turn facilitates the discovery of new information.

Data visualisation is a valuable tool for communicating findings since it is simple to use and facilitates faster innovation and cooperation. Data visualisation technology is becoming more commonplace and spreading across a wide range of fields as a result of the widespread availability of data.

It entails presenting data in a visual way to aid comprehension. It helps making difficult data assessment easy. Bar charts, histograms, graphs, and pie charts are examples.

3.3 DATA MINING

Data mining is the process of breaking down large amounts of raw data into manageable and profitable bits of information. They also look for abnormalities in groups of data and examine the relationships between different data sets in order to find correlations. In many clinical trials, it is primarily utilised to identify behavioural patterns in patient data.

3.4 Machine Learning

ML, which is a subset of AI, is important in data analytics because it involves algorithms that can learn on their own. It allows programmes to take in data and analyse it in order to predict outcomes without the need for someone to expressly configure the system to do so. We can train a machine learning algorithm on a tiny sample of data, and the system will learn as additional data is collected,

improving accuracy over time.

Chapter 4 :LSTM

4.1 DESCRIPTION

LSTM model is used for sequential data. LSTM networks are an unusual type of RNNs that can learn the long term dependencies present in the data-set. RNNs also support data persistence because they have feedback NNs with internal memory.

4.2 INTERNAL MEMORY SPACE OF RNN WHILE PROCESSING NEW DATA

A crucial difference between RNNs and neural networks is that while taking a decision, RNNs considers output as well as the current input, which it has learned from the inputs obtained in earlier situations.

4.3 WHY LSTM?

When traditional RNNs are used they often cause gradient vanishing problems. But LSTM lets us examine long sequences without worrying about such problems. LSTM has 3 gates for each time: Input gate, Output Gate, and forget gate. Forget gate attempts to clear the memory of its previous output.

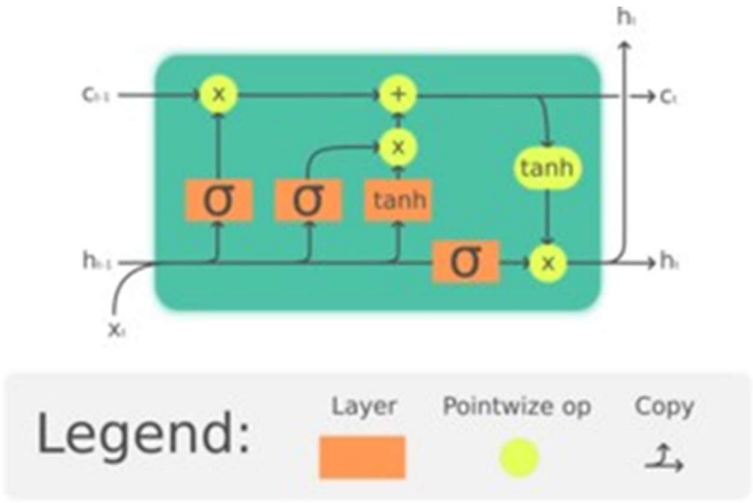


Figure 4.1: LSTM Diagram [20]

$$f_t = \sigma\, W_f \cdot [h_t{-}1, x_t] + b_f$$

Input Gate: This gate is responsible to decide what can be dismissed from the input to tweak the memory.

$$i_t = \sigma\,(W_i \cdot [h_t{-}1, x_t] + b_i)$$

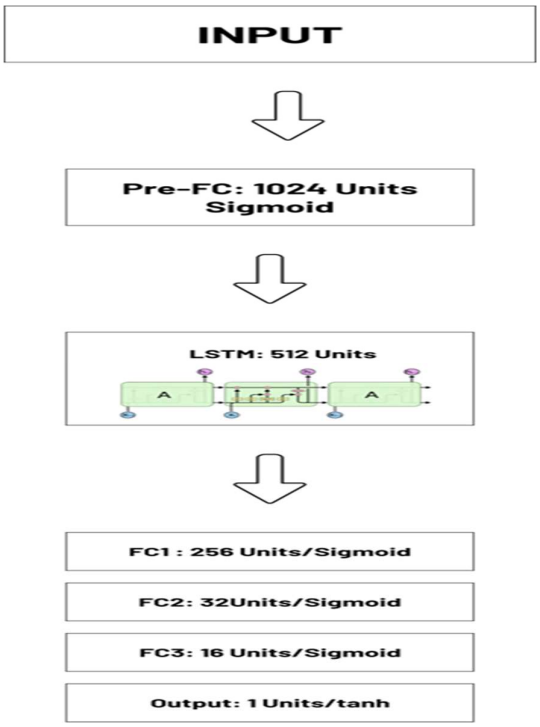
$$Ct = \tanh\,(WC \cdot [h_t{-}1, x_t] + b_C)$$

Output Gate: This gate is responsible to decide the output that can be acquired by

$$o_t = \sigma\,(W_o\,[h_t{-}1, x_t] + b_o)$$

$$t = o_t * \tanh\,(C_t)$$

First, we translate the video clips into a sequence of data points. This data is then further passed via a fully connected layer. This layer has 1024 units and each one of these units uses a sigmoid activation function. Next in this sequence is an LSTM layer. This layer has 512 hidden units. Following these layers are three more fully connected layers until we get the output.



Chapter 5 :BLOCKCHAIN

5.1 INTRODUCTION

A blockchain is a shared database that spans multiple computer network nodes. A blockchain, like a database, stores information in an electronic manner. It is common knowledge that blockchains serve an essential purpose in cryptocurrency systems like Bitcoin by maintaining a public ledger that is both safe and decentralised of all transactions. When it comes to data security and accuracy, blockchain technology is a game changer because it does not necessitate the presence of a trusted third party.

A regular database and a blockchain have quite different ways of organising data. Data is stored in blocks on a blockchain, which are groupings of data. When the maximum storage capacity of a block is achieved, that block is then closed and connected to the block that came before it, creating a chain of data that is referred to as the blockchain. The additional information that follows that freshly added block is assembled into a new block, which is then added to the chain after it is completed.

A database normally organises data into tables, whereas a blockchain, as the name implies, organises data into connected chunks (blocks). This data structure produces an irreversible data timeline whenever it is used in a decentralised implementation. As blocks are completed, they become part of the overall timeline. A time stamp is assigned to each block when it is added to the chain.

The Bottom Line Blockchain was created to serve as the foundation for bitcoin. As it expands and evolves into other sectors of business, including as banking and supply chain management, it has become more open, decentralized, and secure. This technology can inspire creativity for new business models and may be especially valuable for firms that value openness and security.

Small firms, in particular, can benefit from blockchain-based solutions that expedite operations and payments. There are currently a few firms providing this sort of technology, so expect to see many more in the future.

5.2 LAYERED ARCHITECTURE

A Layer-1 network is a blockchain in the decentralized ecosystem, whereas a Layer-2 protocol is a third-party integration that may be used in combination with a Layer-1 blockchain. Layer-1 blockchains include Bitcoin, Litecoin, and Ethereum. Layer-1 scaling solutions boost scalability by supplementing the blockchain protocol's foundation layer. A variety of approaches are now being developed – and used – to directly increase the scalability of blockchain networks.

This is how it works: Layer-1 solutions alter the protocol rules directly in order to boost transaction capacity and speed while supporting additional users and data. Layer-1 scaling options may include increasing the quantity of data in each block or speeding the pace at which blocks are validated in order to boost total network throughput.

5.3 PROOF OF WORK(POW)

A "Proof of Work" is a piece of data that is difficult to develop (expensive, time-consuming), but it is easy for others to verify. PoW-based blockchains rely on specialised computers (referred to as miners) to perform accounting and security responsibilities for the network. Miners are rewarded for their efforts with newly produced coins.

PoW aims to make it prohibitively expensive to attack the network. There is no way to "fake the labour," because changing the ledger requires the use of actual resources. From the standpoint of game theory, Bitcoin's implementation of PoW is a highly elegant balancing of incentives. Incentives motivate humans. Cynthia Dwork and Moni Naor first proposed the concept of PoW in a 1993 article titled "Pricing through Processing or Combating Junk Mail."However, the phrase "Proof of Work" did not become popular until 1999, when Markus Jakobsson and Ari Juels produced a text that formalized PoW:

With Hashcash, Adam Back employed Pow as a technique to avoid email spam. Satoshi Nakamoto, the Bitcoin's pseudonymous developer, later updated and reused the general concept for the Bitcoin Network.

5.4 PROOF OF STAKE(POS)

Sunny King and Scott Nada initially proposed PoS in 2012 as a solution to the problem of high energy costs associated with Bitcoin mining. The first blockchain to adopt PoS was PeerCoin, which

was followed by Blackcoin and NXT.

When learning about PoS, it's helpful to compare mining with staking. Instead of using the energy costs of computational effort to secure the network (mining), nodes in a PoS system give computing resources and "randomly" take turns validating transactions. This keeps the network safe while also reducing energy consumption (staking).

To prevent cheating, each node must stake some tokens, and if they cheat, their tokens will be forfeited. Often referred to as "slashing," it is a way to punish bad actors. These "miners/stakers" are given newly minted tokens as a reward for their contributions to the network, which they support.

Unlike PoW currencies, which can begin with zero coins, a PoS currency must begin with some or all tokens already in existence. This makes widespread distribution of PoS tokens difficult, resulting in wealth and power concentration (oligarchy). This argument of PoS has no bearing on Ethereum, which began as a PoW chain and is only now attempting to convert to PoS. ETH coins were widely distributed during the first 4+ years as a PoW chain.

CHAPTER 6 : MATERIALS AND METHODS

6.1 DATA DESCRIPTION

This dataset includes historical trade data for many crypto assets, including Bitcoin and Ethereum.

Files
train.csv - The training set
timestamp - A timestamp for the minute covered by the row.
Asset_ID - An ID code for the crypto asset.
Count - The number of trades that took place this minute.
Open - The USD price at the beginning of the minute.
High - The highest USD price during the minute.
Low - The lowest USD price during the minute.
Close - The USD price at the end of the minute.
Volume - The number of cryptoasset units traded during the minute.
VWAP - The volume weighted average price for the minute.
Target - 15 minute residualized returns
example_test.csv - An example of the data that will be delivered by the time series API.
example_sample_submission.csv - An example of the data that will be delivered by the time series API. The data is just copied from train.csv.
asset_details.csv - Provides the real name and of the cryptoasset for each Asset_ID and the weight each cryptoasset receives in the metric.

Out[5]:

	timestamp	Asset_ID	Count	Open	High	Low	Close	Volume	VWAP	Target
0	1514764860	2	40.0	2376.5800	2399.5000	2357.1400	2374.5900	19.233005	2373.116392	-0.004218
1	1514764860	0	5.0	8.5300	8.5300	8.5300	8.5300	78.380000	8.530000	-0.014399
2	1514764860	1	229.0	13835.1940	14013.8000	13666.1100	13850.1760	31.550062	13827.062093	-0.014643
3	1514764860	5	32.0	7.6596	7.6596	7.6567	7.6576	6626.713370	7.657713	-0.013922
4	1514764860	7	5.0	25.9200	25.9200	25.8740	25.8770	121.087310	25.891363	-0.008264

Figure 6.1.1 :Train.csv

Out[6]:

	Asset_ID	Weight	Asset_Name
0	2	2.397895	Bitcoin Cash
1	0	4.304065	Binance Coin
2	1	6.779922	Bitcoin
3	5	1.386294	EOS.IO
4	7	2.079442	Ethereum Classic
5	6	5.894403	Ethereum
6	9	2.397895	Litecoin
7	11	1.609438	Monero
8	13	1.791759	TRON
9	12	2.079442	Stellar
10	3	4.406719	Cardano
11	8	1.098612	IOTA
12	10	1.098612	Maker
13	4	3.555348	Dogecoin

Figure 6.1.2 : Asset_details.csv

6.2 EXPLORATORY DATA ANALYSIS

Candlestick chart to show the price of bitcoin last 200 minutes :

Fig-6.2.1

Candlestick chart to show the price of ethereum last 200 minutes :

Fig 6.2.2

Heatmap to show the co-relation between the prices of each cryptocurrencies:

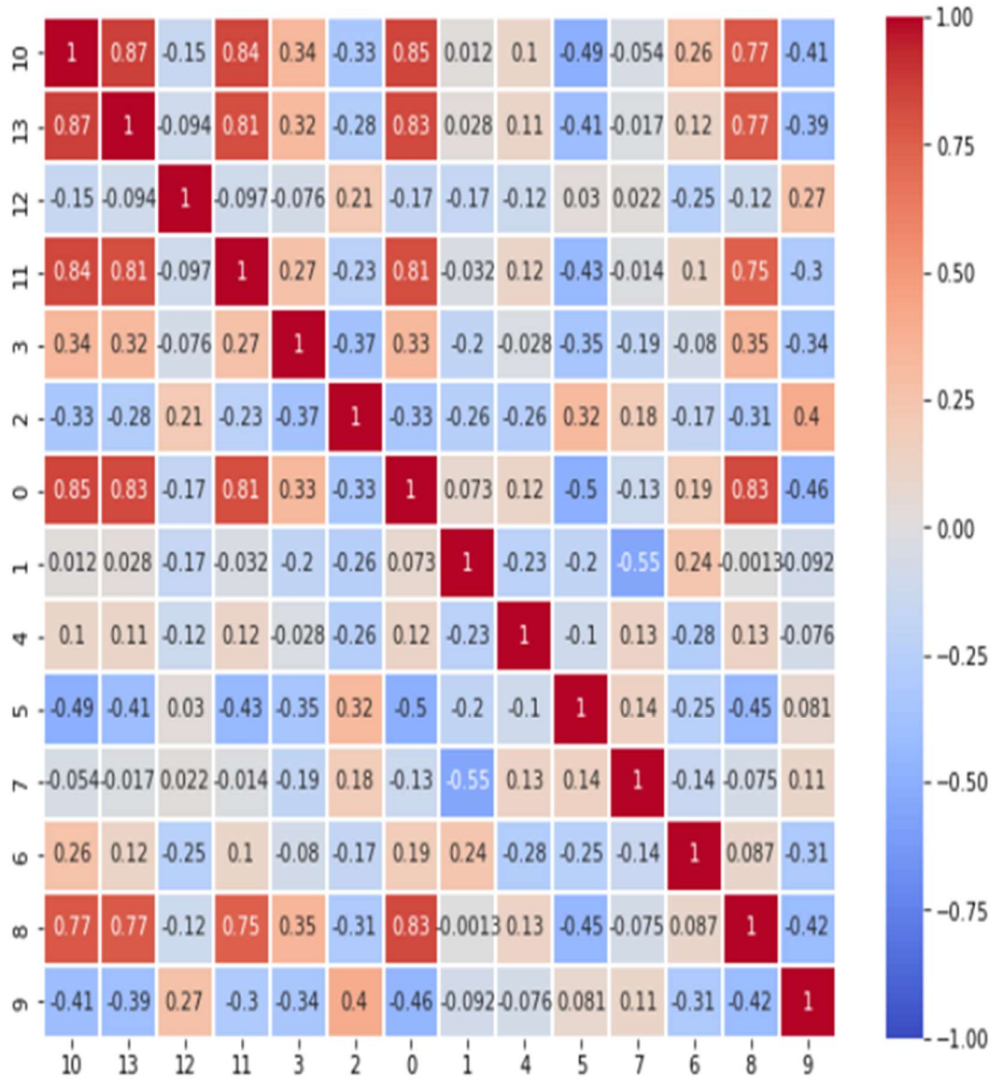
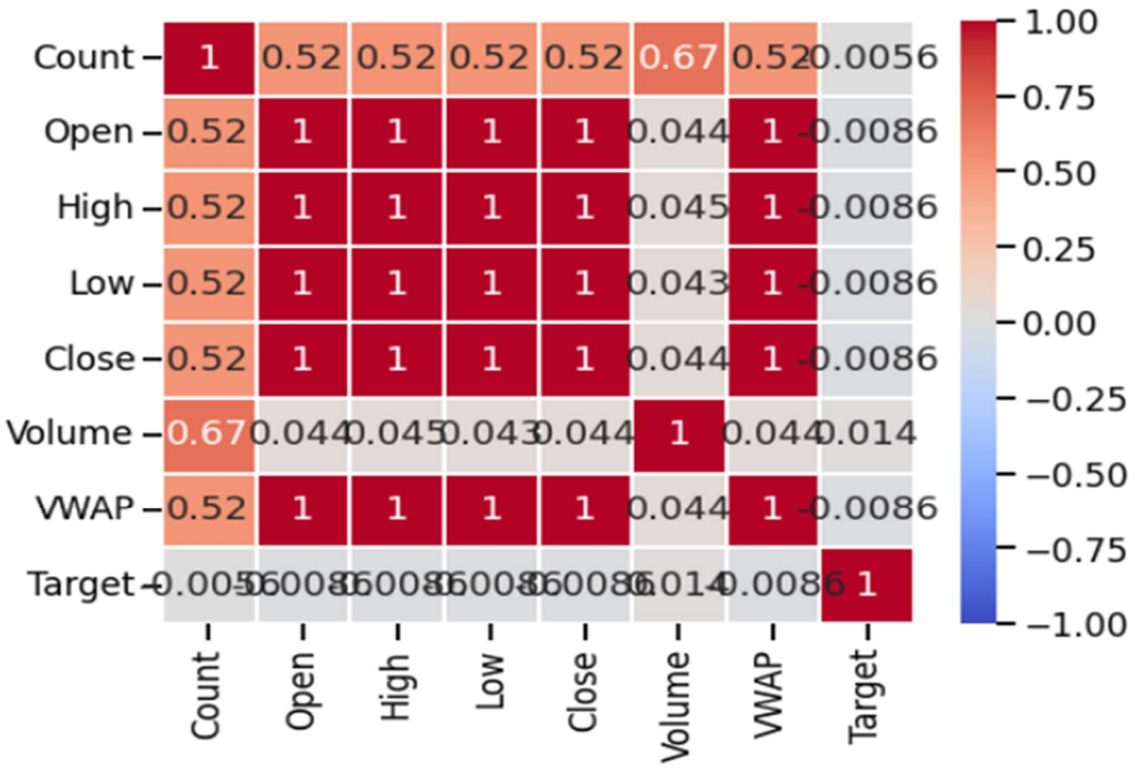


Fig 6.2.3

Heatmap to show the corelation of price of crypto currency to the parameters



p

Fig 6.2.1 illustrates that BTC closing price has decreased and gradually increased in last 200 mins
Fig 6.2.2 illustrates that ETH closing price has been through a sharp fall in the last 200 minutes

Fig 6.2.3 illustrates that the price of cryptocurrencies are heavily dependent on each other

Fig 6.2.4 illustrates that the parameters selected for the crypto forecasting algorithm have a huge effect on the closing prices of respective cryptocurrency.

6.3 MEMORY OPTIMIZATION FUNCTION

The dataset is of 3.5 GB if we feed the given data into the neural network model then it will take a long time to train , so in order to avoid such problems I implemented a memory optimization function in order to reduce the overall size of the data and improve the overall execution time of the algorithm and its efficiency.

In [4]:

```
def reduce_mem_usage(df):
    """ iterate through all the columns of a dataframe and modify the data type
        to reduce memory usage.
    """
    start_mem = df.memory_usage().sum() / 1024**2
    print('Memory usage of dataframe is {:.2f} MB'.format(start_mem))

    for col in df.columns:
        col_type = df[col].dtype

        if col_type != object:
            c_min = df[col].min()
            c_max = df[col].max()
            if str(col_type)[:3] == 'int':
                if c_min > np.iinfo(np.int8).min and c_max < np.iinfo(np.int8).max:
                    df[col] = df[col].astype(np.int8)
                elif c_min > np.iinfo(np.int16).min and c_max < np.iinfo(np.int16).max:
                    df[col] = df[col].astype(np.int16)
                elif c_min > np.iinfo(np.int32).min and c_max < np.iinfo(np.int32).max:
                    df[col] = df[col].astype(np.int32)
                elif c_min > np.iinfo(np.int64).min and c_max < np.iinfo(np.int64).max:
                    df[col] = df[col].astype(np.int64)
            else:
                if c_min > np.finfo(np.float16).min and c_max < np.finfo(np.float16).max:
                    df[col] = df[col].astype(np.float16)
                elif c_min > np.finfo(np.float32).min and c_max < np.finfo(np.float32).max:
                    df[col] = df[col].astype(np.float32)
                else:
                    df[col] = df[col].astype(np.float64)
#         else:
#             df[col] = df[col].astype('category')

    end_mem = df.memory_usage().sum() / 1024**2
    print('Memory usage after optimization is: {:.2f} MB'.format(end_mem))
    print('Decreased by {:.1f}%'.format(100 * (start_mem - end_mem) / start_mem))

    return df
```

Fig 6.3.1 Code Snippet mem_optimize

In [13]:

```
train = pd.read_csv(os.path.join(CFG.INPUT_DIR, 'train.csv')).pipe(reduce_mem_usage)
print(train.shape)
train.head()
```

```
Memory usage of dataframe is 1849.12 MB
Memory usage after optimization is: 716.53 MB
Decreased by 61.2%
(24236806, 10)
```

Fig 6.3.2 Mem_optimize implementation

6.4 EVALUATION METRICS

The evaluation metric is weighted correlation as opposed to a weighted mean of correlation.The metric is defined as follows, where 'a', 'b' and 'weights' are vectors of the same length. 'a' and 'b' are the expected and predicted targets, and ' weights' include the weight of each row, determined by its asset:

```
def weighted_correlation(a, b, weights):

  w = np.ravel(weights)
  a = np.ravel(a)
  b = np.ravel(b)

  sum_w = np.sum(w)
  mean_a = np.sum(a * w) / sum_w
  mean_b = np.sum(b * w) / sum_w
  var_a = np.sum(w * np.square(a - mean_a)) / sum_w
  var_b = np.sum(w * np.square(b - mean_b)) / sum_w

  cov = np.sum((a * b * w)) / np.sum(w) - mean_a * mean_b
  corr = cov / np.sqrt(var_a * var_b)

  return corr
```

Fig 6.4.1 Evaluation Metric

CHAPTER 7: RESULTS

Weighted correlation: The degree to which two values are related or associated is described by their correlation. Values having a strong positive or negative correlation will have a correlation near to 1 or -1, whilst values with no association will have a correlation close to 0.

```
FULL MODEL ****************************************

- Cardano: Validation Score (weighted correlation) = 0.0305

- Bitcoin Cash: Validation Score (weighted correlation) = 0.0058

- Binance Coin: Validation Score (weighted correlation) = 0.0094

- Bitcoin: Validation Score (weighted correlation) = -0.0125

- Dogecoin: Validation Score (weighted correlation) = 0.0418

- EOS.IO: Validation Score (weighted correlation) = -0.0059

- Ethereum Classic: Validation Score (weighted correlation) = 0.0241

- Ethereum: Validation Score (weighted correlation) = 0.0224

- IOTA: Validation Score (weighted correlation) = -0.0037

- Litecoin: Validation Score (weighted correlation) = 0.0067

- Maker: Validation Score (weighted correlation) = 0.0099

- TRON: Validation Score (weighted correlation) = -0.0100

- Stellar: Validation Score (weighted correlation) = 0.0125

- Monero: Validation Score (weighted correlation) = -0.0152
=> Overall Validation Score (weighted correlation) = 0.0172
```

CHAPTER 8 : CONCLUSION

I have learned a lot of things from this internship whether they are technical things or behavioural things which will greatly help me in my future endeavors and will contribute greatly to my success in the future.I have learned a lot about how to work well in a team which has helped me a lot and made me a great team player. I am really grateful to my college for providing me with this opportunity. Machine Learning algorithms like RNN and LGBM may both be useful for forecasting bitcoin values, but the LGBM model excels at finding correlations over longer time horizons. Cryptocurrency maket is growing at great speed and in future will grow at even more faster rate andthe company wishes to offer it's services in the upcoming market and I am very happy to be able to contribute to the company's future suess. However, with such a large variance difficulty, translating this into stunning validation results is challenging. As a consequence, it continues to exist. A challenging assignment It's not the same as overfitting a model and making it incapable of learning. A dropout is a valuable feature for this. Despite the fact that Bayesian optimization was used to aid enhance dropout selection, satisfactory validation results were not achieved.

REFERENCES:

1.Mohammad J. Hamayel, Amani Yousef Owda."A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and bi-LSTM Machine Learning Algorithms", AI, 2021

2. Dongmei Shi, Hongyu Tang. "Research on Safe Driving Evaluation Method Based on Machine Vision Long Short-Term MemoryNetwork", Journal of Electrical and ComputerEngineering, 2021

3. Ranita Biswas, Partha Bhowmick. "On the Functionality and Usefulness of Quadraginta Octants of Naive Sphere", Journal of Mathematical Imaging and Vision, 2017

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	Ram Shayann Goyal		Department	CSE	
Industry/Organization	Praedico Global Research		Date/Duration	10-01-2022 to 12-02-2022	
Criterion	**Poor**	**Average**	**Good**	**Very Good**	**Excellent**
Punctuality/Timely completion of assigned work				✓	
Learning capacity/Knowledge up gradation			✓		
Performance/Quality of work			✓		
Behaviour/Discipline/Team work				✓	
Sincerity/Hard work			✓		
Comment on nature of work done/Area/Topic	Machine Learning Model for Stock Price Prediction.				
OVERALL GRADE (Any one)	**POOR/AVERAGE/GOOD/VERY GOOD/EXCELLENT**				
Name of Industry Mentor	Priyank Gupta				
Signature of Industry Mentor	Priyank				

Receiving Date		**Name of Faculty Mentor**		**Sign**	

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	Ram Bharosa Goyal		Department	CSE	
Industry/Organization	Praedico Global Research PVT Ltd		Date/Duration	13-02-2022 to 28-02-2022	
Criterion	**Poor**	**Average**	**Good**	**Very Good**	**Excellent**
Punctuality/Timely completion of assigned work				✓	
Learning capacity/Knowledge up gradation					✓
Performance/Quality of work					✓
Behaviour/Discipline/Team work					✓
Sincerity/Hard work				✓	
Comment on nature of work done/Area/Topic	Stock Price Prediction Project Machine learning Feature Optimization				
OVERALL GRADE (Any one)	**POOR/AVERAGE/GOOD/VERY GOOD/EXCELLENT** ✓				
Name of Industry Mentor	Priyank Gupta				
Signature of Industry Mentor	Priyank				

Receiving Date		**Name of Faculty Mentor**		**Sign**	

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	Ram Sharavan Goyal		Department	CSE	
Industry/Organization	Praedico Global Research Private Ltd.		Date/Duration	1-03-2021 to 15-03-2021	
Criterion	**Poor**	**Average**	**Good**	**Very Good**	**Excellent**
Punctuality/Timely completion of assigned work					✓
Learning capacity/Knowledge up gradation					✓
Performance/Quality of work				✓	
Behaviour/Discipline/Team work					✓
Sincerity/Hard work					✓
Comment on nature of work done/Area/Topic	Long Short Term Memory Network for stock price prediction (RNN) + Crypto forecasting				
OVERALL GRADE (Any one)	**POOR/AVERAGE/GOOD/VERY GOOD/EXCELLENT** ✓				
Name of Industry Mentor	Priyesh Gupta				
Signature of Industry Mentor	Priyesh				

Receiving Date		**Name of Faculty Mentor**		**Sign**	