A

Major Project

On

BITCOIN PRICE PREDICTION AND FORECASTING

(Submitted in partial fulfillment of the requirements for the award of Degree)

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

By

KIRAN LENKA	(187R1A05A2)
NADAKUDITI DEVI	(187R1A05A0)
D.V.RITHIKA	(187R1A05B4)

Under the Guidance of

RANJITH REDDY K

(Assistant Professor)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CMR TECHNICAL CAMPUS

UGC AUTONOMOUS

(Accredited by NAAC, NBA, Permanently Affiliated to JNTUH, Approved by AICTE, New Delhi)
Recognized Under Section 2(f) & 12(B) of the UGC Act.1956,
Kandlakoya (V), Medchal Road, Hyderabad-501401.

2018-2022

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the project entitled "**BITCOIN PRICE PREDICTION AND FORECASTING**", being submitted by **KIRAN LENKA (187R1A05A2)**, **NADAKUDITI DEVI (187R1A05A0)**, **D.V.RITHIKA (187R1A05B4)** in partial fulfillment of the requirements for the award of the degree of B. Tech in Computer Science and Engineering of the Jawaharlal Nehru Technological University Hyderabad, is a record of bonafide work carried out by him/her under our guidance and supervision during the year 2021-2022.

The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

RANJITH REDDY K
(Assistant Professor)
INTERNAL GUIDE

Dr. A. RAJI REDDY
DIRECTOR

Dr. K. Srujan Raju
HOD

EXTERNAL EXAMINER

Submitted for viva voce Examination held on _____

ACKNOWLEDGEMENT

Apart from the efforts of us, the success of any project depends largely on the encouragement and guidelines of many others. We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project.

We take this opportunity to express our profound gratitude and deep regard to our guide **Mr. Ranjith Reddy K,** Assistant Professor, Department of CSE for his exemplary guidance, monitoring and constant encouragement throughout the project work. The blessing, help and guidance given by him shall carry us a long way in the journey of life on which we are about to embark.

We also take this opportunity to express a deep sense of gratitude to Project Review Committee (PRC) **Mr. A. Uday Kiran , Mr. J. Narasimha Rao, Dr. T. S. Mastan Rao, Mrs. G. Latha , Mr. A. Kiran Kumar,** for their cordial support, valuable information and guidance, which helped us in completing this task through various stages.

We are also thankful to **Dr. K. Srujan Raju,** Head, Department of Computer Science and Engineering for providing encouragement and support for completing this project successfully.

We are obliged to **Dr. A. Raji Reddy,** Director for being cooperative throughout the course of this project. We would like to express our sincere gratitude to Sri. **Ch. Gopal Reddy,** Chairman for providing excellent infrastructure and a nice atmosphere throughout the course of this project.

The guidance and support received from all the members of **CMR Technical Campus** who contributed to the completion of the project. We are grateful for their constant support and help.

Finally, we would like to take this opportunity to thank our family for their constant encouragement, without which this assignment would not be completed. We sincerely acknowledge and thank all those who gave support directly and indirectly in the completion of this project.

KIRAN LENKA (187R1A05A2)
NADAKUDITI DEVI (187R1A05A0)
D.V.RITHIKA (187R1A05B4)

ABSTRACT

Cryptocurrency, especially Bitcoin, is one of the most volatile markets today and has gained a lot of attention from investors across the globe. Cryptocurrency, being a novel technique for transaction system, has led to a lot of confusion among the investors and any rumors or news on social media has been claimed to significantly affect the prices of cryptocurrencies. Bitcoin is currently the leading global provider of cryptocurrency. Cryptocurrency allows users to safely and anonymously use the Internet to perform digital currency transfers and storage. In recent years, the Bitcoin network has attracted investors, businesses, and corporations while facilitating services and product deals. Moreover, Bitcoin has made itself the dominant source of decentralized cryptocurrency. Bitcoin is one of the oldest and biggest cryptocurrencies being traded as of now, in terms of the volume being traded. It is a digital asset over which central banks/financial Institutions have no control or regulations.

Over the past 2 years, Bitcoin has seen its highest price around $20000 and its lowest price around $900. It is very sporadic and this is one of the most important reasons which attracted us to analyse and predict its price. Bitcoin has a market share of more than 55% as compared to other cryptocurrencies, being followed by Ethereum at 8.57% .It is very sporadic and this is one of the most important reasons which attracted us to analyze and predict its price. Here, we intend to study the prediction of Bitcoin prices using Machine Learning Techniques and prepare a strategy to maximize gains for investors. Proper Bitcoin prediction model will help the investors beforehand when investing in any cryptocurrency and help them reduce the chances of facing a loss by using various machine learning techniques combined with LSTM model and performing sentiment analysis on tweets, Reddit posts and news headlines to understand the situation around the cryptocurrency.

LIST OF FIGURES

LIST OF SCREENSHOTS

LIST OF TABLES

TABLE OF CONTENTS

1. INTRODUCTION

1. INTRODUCTION

Stock market is one of the most volatile data available in terms of Machine learning datasets. Researchers have been long trying to predict the stock market and any breakthrough in this field would result in, literally, the people being able to mint money. Cryptocurrencies, to be specific, have gained a lot of traction in recent years from investors across the globe. Cryptocurrency being a novel technique for transaction systems has led to a lot of confusion among investors and any rumours or news on social media has been claimed to significantly affect the prices of cryptocurrencies. Cryptocurrency, especially Bitcoin, is one of the most volatile markets today and has gained a lot of attention from investors across the globe. Bitcoin is one of the oldest and biggest cryptocurrencies being traded as of now, in terms of the volume being traded. It is a digital asset over which central banks or any financial Institutions have no control or regulations. Bitcoin has a market share of more than 55% as compared to other cryptocurrencies, being followed by Ethereum at 8.57%. It is very sporadic and this is one of the most important reasons which attracted us to analyze and predict its price.

The goal of this study is to predict prices for Bitcoin using Machine Learning Techniques for the next day and prepare a strategy to maximize gains for investors. We also aim to find out if there is a co-relation between fluctuating Bitcoin prices and related news.

1.1 PROJECT PURPOSE

The idea of this project is to use Deep learning techniques such as LSTM model combined with Sentiment analysis and prediction using Machine Learning, With the help of this prediction model, investors will be able to know the changes in prices of Bitcoin beforehand and it will also help them reduce the chances of facing a loss. This system with little modification can be used to predict the price of any cryptocurrency in the near future.

1.2 PROJECT SCOPE

This project is titled “Bitcoin Price Prediction and Forecasting”. Sentiment analysis and prediction using a Machine Learning tool that analyzes texts for polarity, from positive to negative will be performed and with the examples of emotions in text, machines automatically learn how to detect sentiment without human input. Sentiment analysis on Reddit and Twitter are popular social media platforms where discussion on CryptoCurrency is higher when compared to other sites. Using Visual elements like charts, graphs, maps to understand and see trends, outliers, and patterns in data for the graphical representation. Training and testing of dataset extracted from CryptoCompare tool to predict the price of Bitcoin using the LSTM neural network along with Sentiment analysis to analyze the model performance.

1.3 PROJECT FEATURES

- Deep learning techniques such as the LSTM model combined with Sentiment analysis and prediction using Machine Learning, will help them reduce the chances of facing a loss.
- Sentiment analysis and market sentiment prediction on popular social media platforms such as Twitter and Reddit using tweets and posts related to Bitcoin will be performed.
- With the help of this prediction model, investors will be able to know the variation in the prices of bitcoin and invest in it carefully.

2. SYSTEM ANALYSIS

2. SYSTEM ANALYSIS

SYSTEM ANALYSIS

System Analysis is the important phase in the system development process. The System is studied to the minute details and analyzed. The system analyst plays an important role of an interrogator and dwells deep into the working of the present system. In analysis, a detailed study of these operations performed by the system and their relationships within and outside the system is done. A key question considered here is, “what must be done to solve the problem?” The system is viewed as a whole and the inputs to the system are identified. Once analysis is completed the analyst has a firm understanding of what is to be done.

2.1 PROBLEM DEFINITION

Given the vastness of investments in Bitcoin, Cryptocurrencies are relatively unpredictable compared to traditional financial predictions like stock market prediction. Their high volatility leads to the great potential of high profit if intelligent investing strategies are taken. Proper Bitcoin prediction model will help the investors beforehand when investing in any cryptocurrency and help them reduce the chances of facing a loss by using various machine learning techniques combined with LSTM model and performing sentiment analysis on tweets, Reddit posts and news headlines to understand the situation around the cryptocurrency.

2.2 EXISTING SYSTEM

The value of Bitcoins is constantly fluctuating according to demand. Many machine learning techniques have been explored for stock price prediction (Ballings et al. 2015). Almeida et al. (2015) reviewed an artificial neural network (ANN) model to predict the Bitcoin price using the last day price and turnover volumes. The main problem with their method is the requirement of a large amount of data for the

prediction. ANN and Support Vector Regression (SVR) are two widely used machine learning algorithms for predicting stock price and stock market index values (Patel et al. 2015).

2.2.1 LIMITATIONS OF EXISTING SYSTEM

- The existing works consider a lot of features that are less useful in the prediction of stock prices.
- Some works fail to recognize many features that provide insights into the data.
- Existing works are mostly suitable for relatively less varying stock market prices, and cannot be applied for highly fluctuating cryptocurrencies such as Bitcoin.

2.3 PROPOSED SYSTEM

Sentiment analysis and market sentiment prediction on popular social media platforms such as Twitter and Reddit using tweets and posts related to Bitcoin will be performed. Bitcoin price data collected using the CryptoCompare API (provides highly reliable historical data for the public) and Data visualization will help investors to know the volatility in the prices of bitcoins. We aim to train and test the dataset to predict the price of Bitcoin using the LSTM neural network. Visualize the prediction results in the form of graphical representation that contains the information and the data.

2.3.1 ADVANTAGES OF THE PROPOSED SYSTEM

- We will be able to predict the sentiment of tweets, news headlines and Reddit posts related to bitcoin.
- With the help of this system, one can predict the stock price of Bitcoin and make their investments carefully.
- The risk of loss can be avoided by predicting the price of Bitcoin beforehand.

- This system with little modification can be used to predict the price of any cryptocurrency.

2.4 FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and a business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. Three key considerations involved in the feasibility analysis are

- Economic Feasibility
- Technical Feasibility
- Social Feasibility

2.4.1 TECHNICAL FEASIBILITY

To determine whether the proposed system is technically feasible, we should take into consideration the technical issues involved behind the system. This Application uses the Machine learning techniques, which are rampantly employed these days worldwide. The world without Artificial Intelligence is incomprehensible today that goes to the proposed system is technically feasible.

2.4.2 ECONOMICAL FEASIBILITY

To decide whether a project is economically feasible, we have to consider various factors as:

- Cost benefit analysis
- Long-term returns
- Maintenance cost

It requires average computing capabilities and access to the internet, which are very basic requirements and can be afforded by any organization hence it doesn’t

incur additional economic overheads, which renders the system economically feasible.

2.4.3 BEHAVIOURAL FEASIBILITY

This includes the following questions: Is there sufficient support for the users? Will the proposed system cause harm? The project would be beneficial because it satisfies the objectives when developed and installed. All behavioural aspects are considered carefully and conclude that the project is behaviourally feasible.

2.5 HARDWARE AND SOFTWARE REQUIREMENTS

2.5.1 HARDWARE REQUIREMENTS

Hardware interfaces specify the logical characteristics of each interface between the software product and the hardware components of the system. The following are some hardware requirements.

- Ram : 8GB Ram and above
- Hard Disk : 20GB and above
- Processor : Intel(R)Core(™)i5-8250U CPU
- GPU : AMD Radeon 530 Graphics

2.5.2 SOFTWARE REQUIREMENTS

Software Requirements specifies the logical characteristics of each interface and software components of the system. The following are some software requirement

- Operating system : Windows/Linux
- Programming Language : Python
- Tools : Google Colab, CryptoCompare
- Libraries : NumPy, Matplotlib, SKLEARN, Tweepy, PRAW, TensorFlow.

3. ARCHITECTURE

3. ARCHITECTURE

3.1 PROJECT ARCHITECTURE

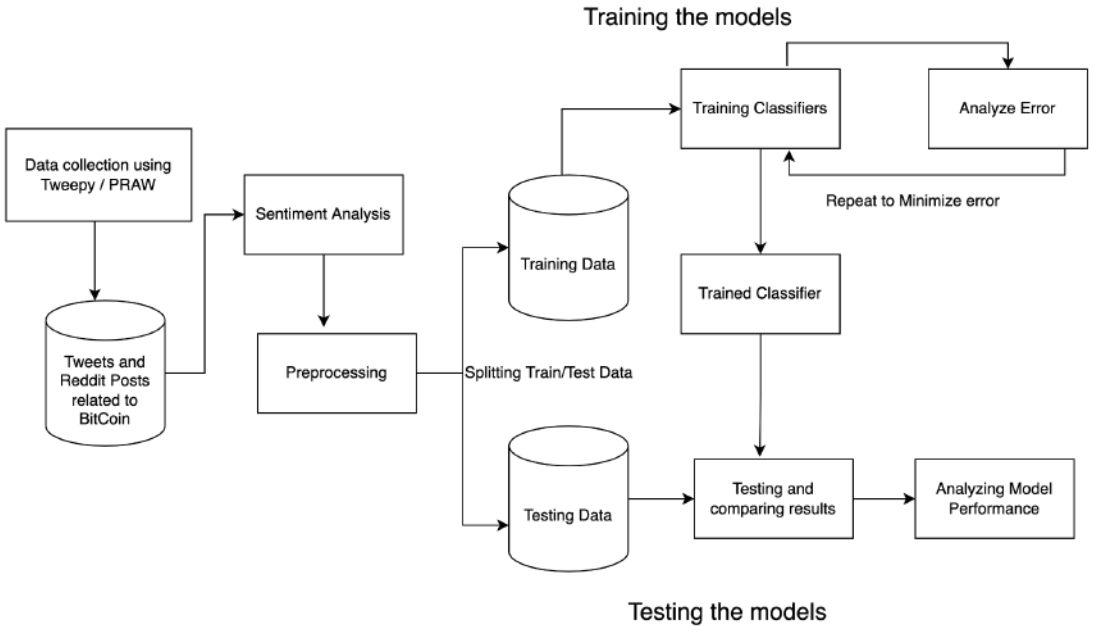


Figure 3.1 Project architecture of Market Sentiment Prediction

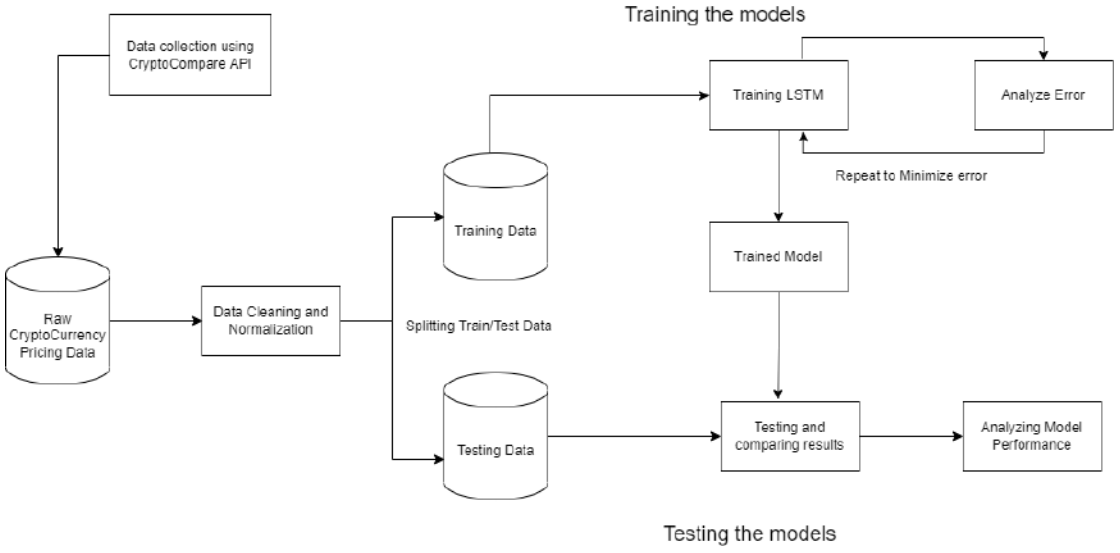


Figure 3.2 Project architecture of BTC Price Prediction

3.2 DESCRIPTION

Input Data: In this system there are two types of input data. The first type of input data is the tweets of current trends and prices related to Bitcoin. It is a csv file with approximately 8000 tweets. The second type of input data is the Reddit posts related to Bitcoin prices where content is socially curated and promoted by the social website members through upvoting. This is used to build and train the models and to evaluate their performance.

Data Cleaning and Preprocessing: In this step, the input data is cleaned and preprocessed. This step involves the removal of emoticons, hashtags, mentions, numerical values, punctuations and stop words, stemming and lemmatization.

Sentiment Analysis and Classification: Here the machine visualizes how many positive, neutral and negative tweets are present using bar plots, training and testing 6 different models. It then plots the performance between these models.

Price Prediction: In this step, visualizes price changes in history, compare with Nifty-50 and BSE-Sensex, view volatility, collect data using CryptoCompare, train and test LSTM model and view performance.

Building Models: Multiple machine learning algorithms are built based on the features extracted using various natural language processing techniques. Such as Logistic Regression, Random Forest, SGD, Support-Vector Machine, LightGBM, Gradient Boosting Classifier.

Evaluating Models: The models built are carefully evaluated on the test dataset using various performance metrics such as Precision, Accuracy, F1 Score etc. and are

compared. The whole system can be divided into two major phases such as training and testing.

Training the models: Here multiple models are trained using the labelled dataset and this dataset is used to train Machine Learning Algorithms and it is fed with sufficient training data to learn from.

Testing the models: The trained models are tested and their results are considered to fine tune the parameters of the model. It is the process where the performance of a fully trained model is evaluated on a testing set.

3.3 DATA FLOW DIAGRAM

3.3.1 DATA FLOW DIAGRAM FOR BTC PRICE PREDICTION

The flow of the diagram for BTC price prediction in this system demonstrates how the raw data from cryptocompare is fed into the system. The data is cleaned and then divided into two sets, one for training and the other for testing. We also apply a variety of algorithms to it. The trained models are put to the test, and the results are used to fine-tune the model's parameters.

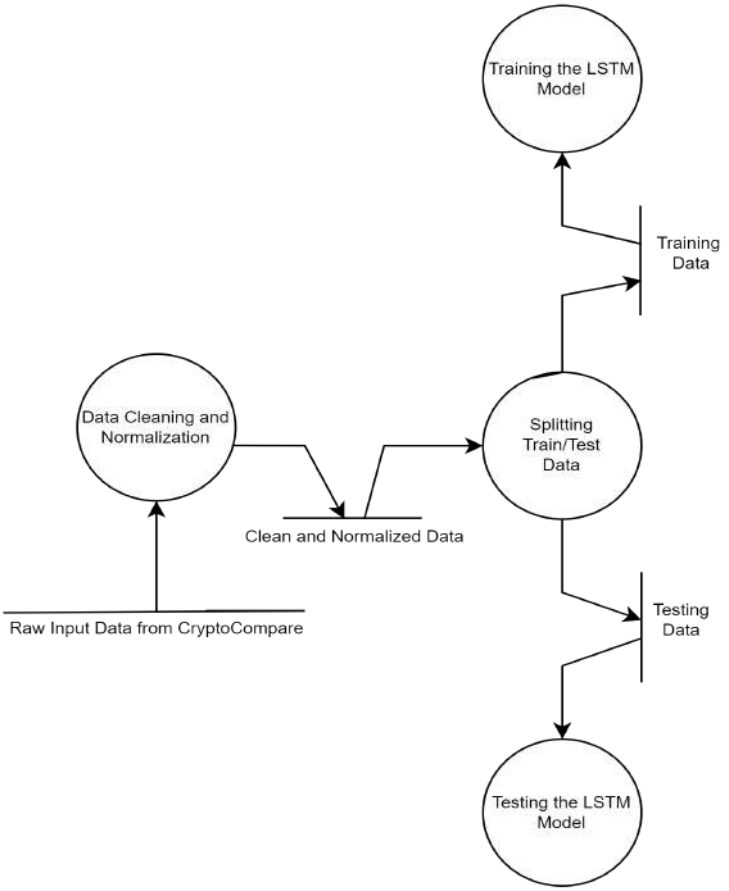


Figure 3.3 Data Flow Diagram for BTC Price Prediction using LSTM

3.3.2 DATA FLOW DIAGRAM FOR MARKET SENTIMENT ANALYSIS

In this system, the flow of the diagram for Market sentiment analysis shows how the data is first collected from Tweepy and PRAW, sentiment analysis is performed on it. The data, has a sentiment score based on the positive, neutral, negative tweets or reddit posts. Data is pre-processed and then it is split into training data and testing data where we apply various classification algorithms to check the results of this project which will help us predict the trends beforehand when investing in any cryptocurrency.

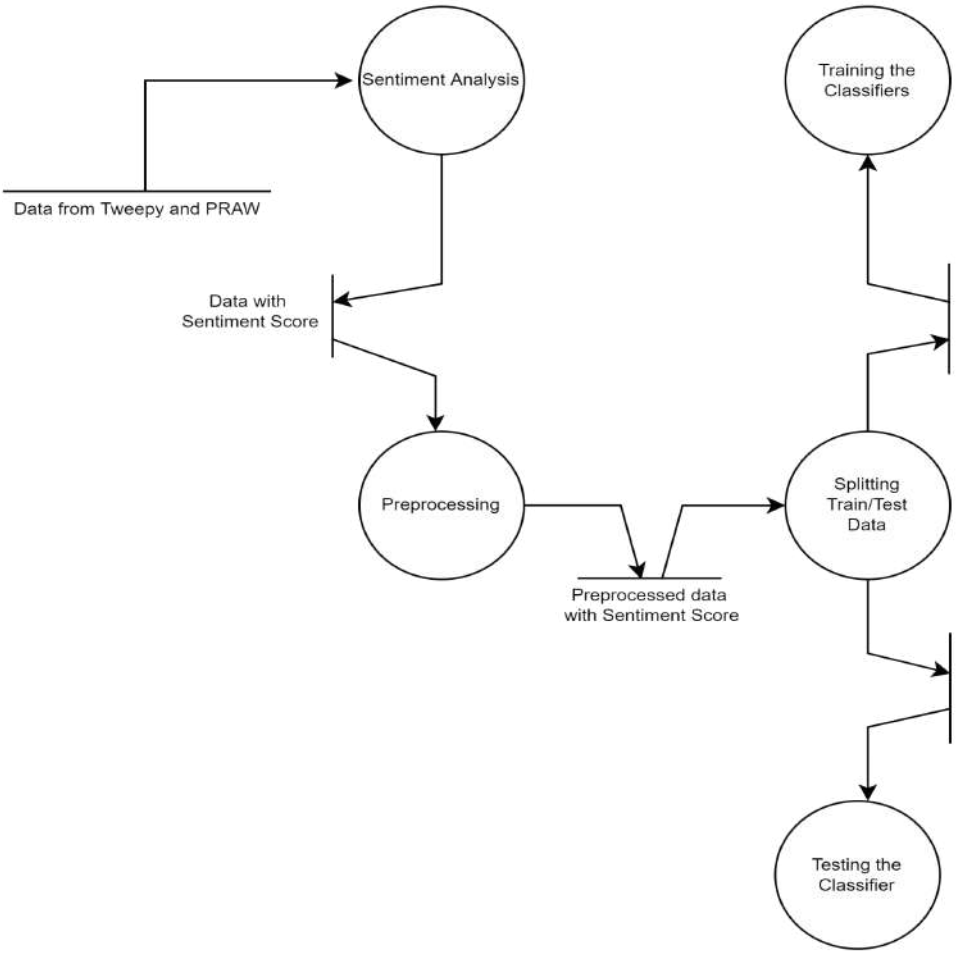


Figure 3.4 Data Flow Diagram for Bitcoin Market Sentiment Analysis using data from Twitter and Reddit

4. IMPLEMENTATION

4. IMPLEMENTATION

4.1 SAMPLE CODE:

```
# %tensorflow_version 2.x
import json
import requests
from keras.models import Sequential
from keras.layers import Activation, Dense, Dropout, LSTM
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import seaborn as sns
from sklearn.metrics import mean_absolute_error
# %matplotlib inline

endpoint = 'https://min-api.cryptocompare.com/data/histoday'
res = requests.get(endpoint +'?fsym=BTC tsym=USD limit=1000')
#print (res.content['Data'])
hist = pd.DataFrame(json.loads(res.content)['Data'])
hist = hist.set_index('time')
hist.index = pd.to_datetime(hist.index, unit='s')
target_col = 'close'

hist.drop(["conversionType", "conversionSymbol"], axis =
'columns', inplace = True)

hist.tail(5)

def train_test_split(df, test_size=0.2):
split_row = len(df) - int(test_size * len(df))
train_data = df.iloc[:split_row]
test_data = df.iloc[split_row:]
return train_data, test_data
```

```
train, test = train_test_split(hist, test_size=0.2)
def line_plot(line1, line2, label1=None, label2=None,
title='', lw=2):
fig, ax = plt.subplots(1, figsize=(13, 7))
ax.plot(line1, label=label1, linewidth=lw)
ax.plot(line2, label=label2, linewidth=lw)
ax.set_ylabel('price [CAD]', fontsize=14)
ax.set_title(title, fontsize=16)
ax.legend(loc='best', fontsize=16);

line_plot(train[target_col], test[target_col], 'training',
'test', title='')

def normalise_zero_base(df):
return df / df.iloc[0] - 1

def normalise_min_max(df):
return (df - df.min()) / (data.max() - df.min())

def extract_window_data(df, window_len=5, zero_base=True):
window_data = []
for idx in range(len(df) - window_len):
tmp = df[idx: (idx + window_len)].copy()
if zero_base:tmp = normalise_zero_base(tmp)
window_data.append(tmp.values)
return np.array(window_data)

def prepare_data(df, target_col, window_len=10,
zero_base=True, test_size=0.2):

train_data, test_data = train_test_split(df,
test_size=test_size)
X_train = extract_window_data(train_data, window_len,
zero_base)
X_test = extract_window_data(test_data, window_len,
zero_base)
y_train = train_data[target_col][window_len:].values
y_test = test_data[target_col][window_len:].values
if zero_base:
y_train = y_train /
train_data[target_col][:-window_len].values - 1
```

```
y_test = y_test / test_data[target_col][:-window_len].values
- 1
return train_data, test_data, X_train, X_test, y_train,
y_test

def build_lstm_model(input_data, output_size, neurons=100,
activ_func='linear', dropout=0.2, loss='mse',
optimizer='adam'):
model = Sequential()
model.add(LSTM(neurons, input_shape=(input_data.shape[1],
input_data.shape[2])))
model.add(Dropout(dropout))
model.add(Dense(units=output_size))
model.add(Activation(activ_func))
model.compile(loss=loss, optimizer=optimizer)
return model

np.random.seed(42)
window_len = 5
test_size = 0.2
zero_base = True
lstm_neurons = 100
epochs = 20
batch_size = 32
loss = 'mse'
dropout = 0.2
optimizer = 'adam'

train, test, X_train, X_test, y_train, y_test = prepare_data(
hist, target_col, window_len=window_len, zero_base=zero_base,
test_size=test_size)

model = build_lstm_model(X_train, output_size=1,
neurons=lstm_neurons, dropout=dropout, loss=loss,
optimizer=optimizer)
history = model.fit(X_train, y_train,
validation_data=(X_test, y_test), epochs=epochs,
batch_size=batch_size, verbose=1, shuffle=True)
```

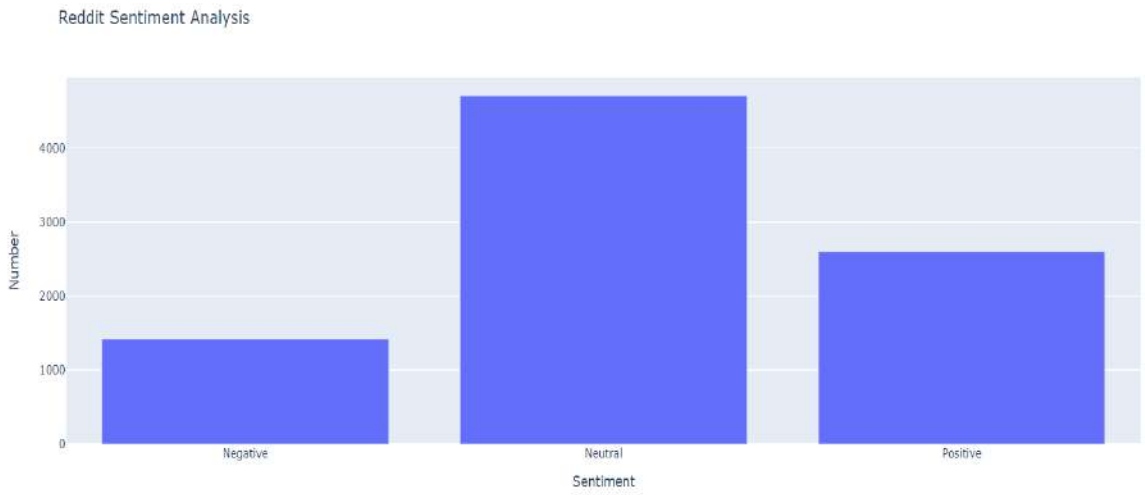
```
import matplotlib.pyplot as plt
plt.plot(history.history['loss'],'r',linewidth=2,
label='Train loss')
plt.plot(history.history['val_loss'], 'g',linewidth=2,
label='Validation loss')
plt.title('LSTM')
plt.xlabel('Epochs')
plt.ylabel('MSE')
plt.show()

targets = test[target_col][window_len:]
preds = model.predict(X_test).squeeze()
mean_absolute_error(preds, y_test)

from sklearn.metrics import mean_squared_error
MAE=mean_squared_error(preds, y_test)
MAE

from sklearn.metrics import r2_score
R2=r2_score(y_test, preds)
R2

preds = test[target_col].values[:-window_len] * (preds + 1)
preds = pd.Series(index=targets.index, data=preds)
line_plot(targets, preds, 'actual', 'prediction', lw=3)
```

5. SCREENSHOTS

5. SCREENSHOTS

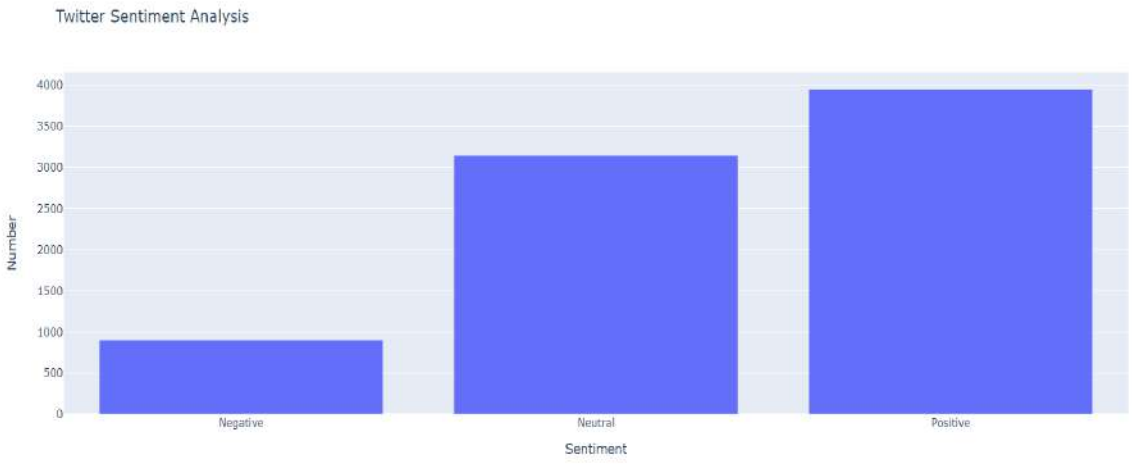
5.1 SENTIMENT ANALYSIS

5.1.1 REDDIT SENTIMENT ANALYSIS



Screenshot 5.1 Reddit Sentiment Analysis

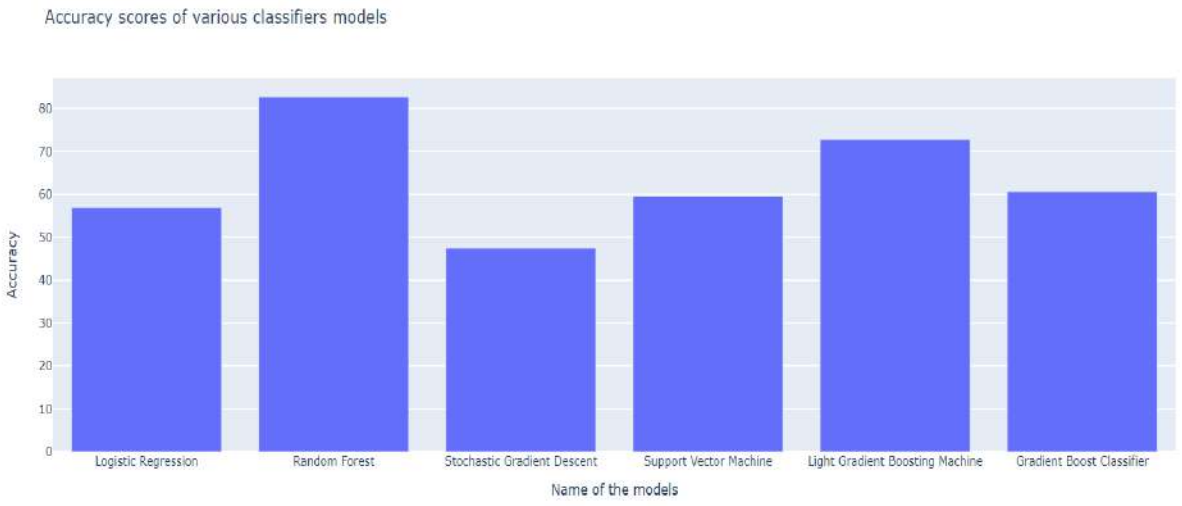
5.1.2 TWITTER SENTIMENT ANALYSIS



Screenshot 5.2 Twitter Sentiment Analysis

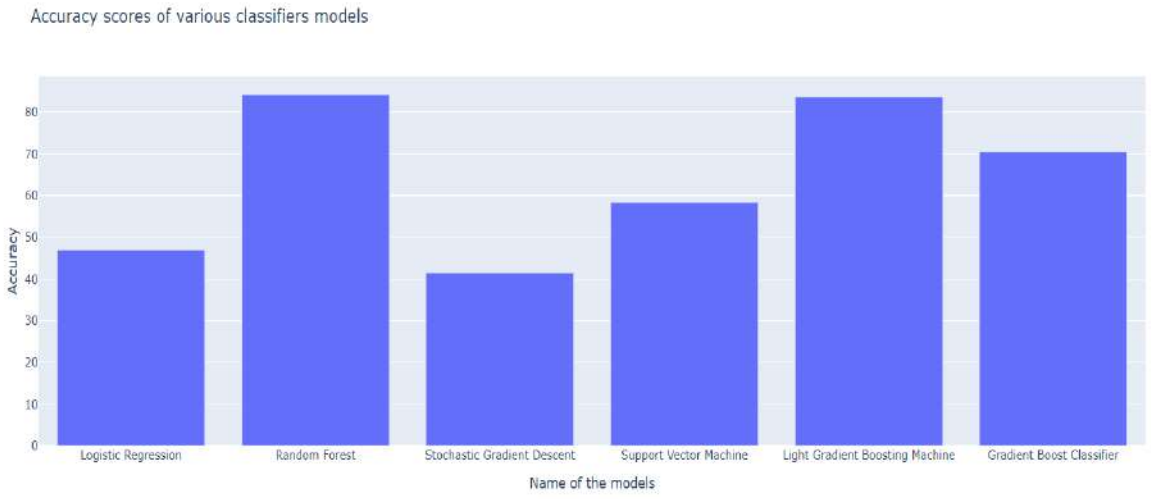
5.2 PERFORMANCE COMPARISON BETWEEN DIFFERENT CLASSIFIERS

5.2.1 REDDIT POSTS



Screenshot 5.3 Accuracy Score of various Classifier model for the Reddit posts

5.2.2 TWITTER POSTS



Screenshot 5.4 Accuracy Score of various Classifier model for the Twitter posts

5.3 REAL-TIME CRYPTOCURRENCY DATA

```
[4]  hist.tail(5)
```

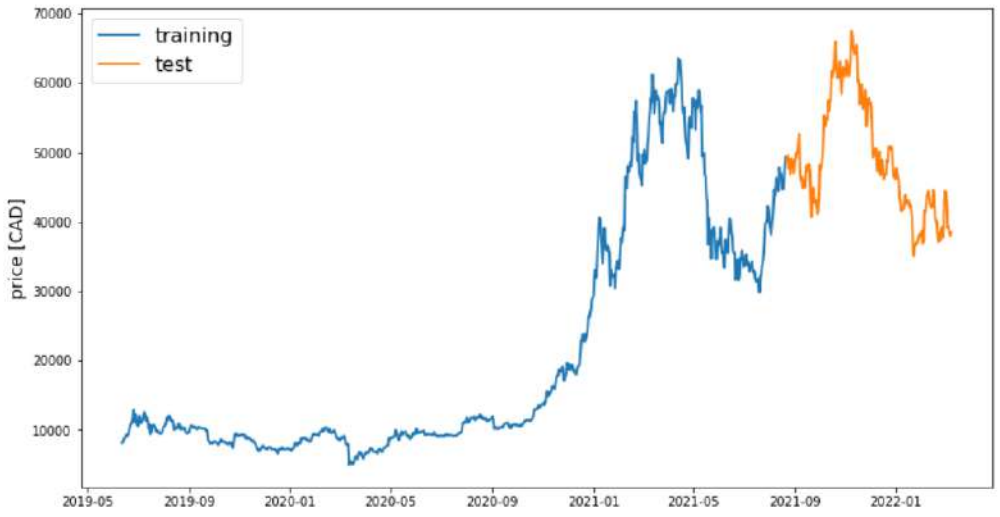
time	high	low	open	volumefrom	volumeto	close
2022-03-04	42520.95	38621.24	42474.37	50126.39	2.043288e+09	39154.40
2022-03-05	39596.42	38652.98	39154.40	15938.79	6.242895e+08	39407.02
2022-03-06	39686.74	38133.84	39407.02	21034.62	8.183791e+08	38428.92
2022-03-07	39536.14	37191.99	38428.92	40474.56	1.547870e+09	38030.63
2022-03-08	38922.20	37905.43	38030.63	12762.67	4.892734e+08	38526.38

Screenshot 5.5 Cryptocurrency prices using all the trading features

5.4 CRYPTOCURRENCY PRICES

5.4.1 CRYPTOCURRENCY PRICES IN INDIAN RUPEES AS A FUNCTION OF TIME

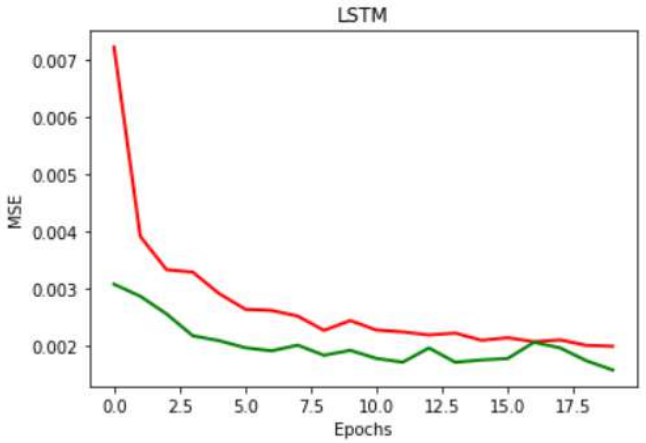
```
line_plot(train[target_col], test[target_col], 'training', 'test', title='')
```



Screenshot 5.6 Cryptocurrency prices in Indian rupees

5.4.2 GRAPH REPRESENTING MSME AND EPOCHS USING LSTM

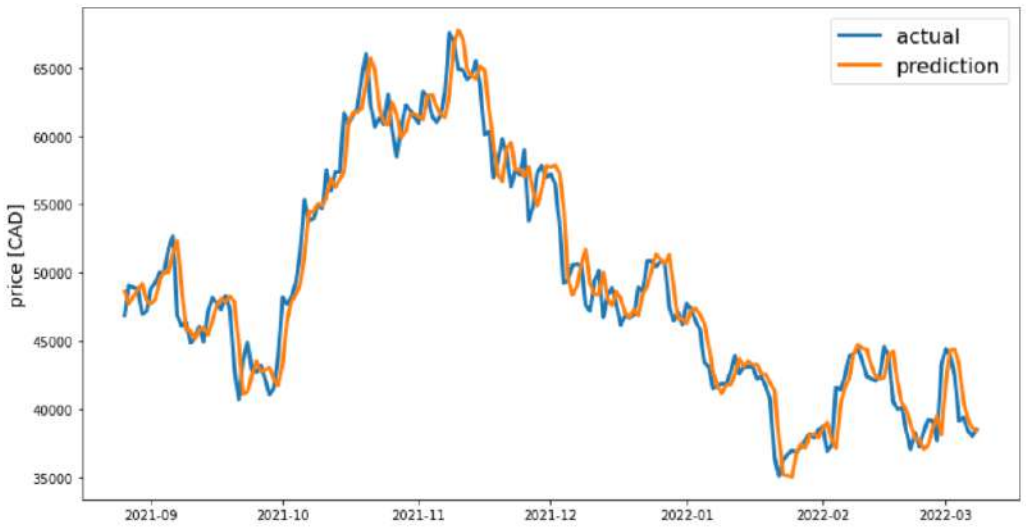
```
[16] import matplotlib.pyplot as plt
     plt.plot(history.history['loss'],'r',linewidth=2, label='Train loss')
     plt.plot(history.history['val_loss'], 'g',linewidth=2, label='Validation loss')
     plt.title('LSTM')
     plt.xlabel('Epochs')
     plt.ylabel('MSE')
     plt.show()
```



Screenshot 5.7 Graphical representation of LSTM

5.4.3 GRAPH PLOT OF ACTUAL AND PREDICTED PRICES

```
preds = test[target_col].values[:-window_len] * (preds + 1)
preds = pd.Series(index=targets.index, data=preds)
line_plot(targets, preds, 'actual', 'prediction', lw=3)
```



Screenshot 5.8 Prediction of Cryptocurrency prices in Real time

6. TESTING

6. TESTING

6.1 SOFTWARE TESTING

Software testing is a critical element of software quality assurance and represents the ultimate review of specification, design and code.

6.1.1 TESTING OBJECTIVES

- To ensure that during operation the system will perform as per specification. To make sure that system meets the user requirements during operation
- To make sure that during the operation, incorrect input, processing and output will be detected
- To see that when correct inputs are fed to the system the outputs are correct
- To verify that the controls incorporated in the same system as intended
- Testing is a process of executing a program with the intent of finding an error
- A good test case is one that has a high probability of finding an as yet undiscovered error

The software developed has been tested successfully using the following testing strategies and any errors that are encountered are corrected and again the part of the program or the procedure or function is put to testing until all the errors are removed. A successful test is one that uncovers an as yet undiscovered error. Note that the result of the system testing will prove that the system is working correctly. It will give confidence to system designer, users of the system, prevent frustration during the implementation process etc.

6.2 TEST CASE DESIGN:

6.2.1 WHITE BOX TESTING

White box testing is a testing case design method that uses the control structure of the procedure design to derive test cases. All independent paths in a

module are exercised at least once, all logical decisions are exercised at once, execute all loops at boundaries and within their operational bounds exercise internal data structure to ensure their validity. Here the customer is given three chances to enter a valid choice out of the given menu. After which the control exits the current menu.

6.2.2 BLACK BOX TESTING

Black Box Testing attempts to find errors in following areas or categories, incorrect or missing functions, interface error, errors in data structures, performance error and initialization and termination error. Here all the input data must match the data type to become a valid entry. The following are the different tests at various levels:

6.2.3 UNIT TESTING

Unit testing is essentially for the verification of the code produced during the coding phase and the goal is to test the internal logic of the module/program. In the Generic code project, the unit testing is done during the coding phase of data entry forms whether the functions are working properly or not. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application. It is done after the completion of an individual unit before integration. This is a structural testing that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

6.2.4 INTEGRATION TESTING

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfactory, as shown by successfully unit testing, the combination of components is correct and consistent. Integration

testing is specifically aimed at exposing the problems that arise from the combination of components.

6.2.5 VALIDATION TESTING

This testing concentrates on confirming that the software is error-free in all respects. All the specified validations are verified and the software is subjected to hard-core testing. It also aims at determining the degree of deviation that exists in the software designed from the specification; they are listed out and are corrected.

6.2.6 FUNCTIONAL TESTING

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input	: identified classes of valid input must be accepted.
Invalid Input	: identified classes of invalid input must be rejected.
Functions	: identified functions must be exercised.
Output	: identified classes of application outputs must be exercised.
Systems/Procedures	: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identifying Business process flows; data fields, predefined processes.

6.2.7 SYSTEM TESTING

This testing is a series of different tests whose primary is to fully exercise the computer-based system. This involves:

- Implementing the system in a simulated production environment and testing it.
- Introducing errors and testing for error handling.

6.3 TEST CASES

6.3.1 TEST CASE 1:

Test case for Training data:

The model is trained using Tweets, Reddit posts and News headlines related to the cryptocurrency such as Bitcoin.

Test case ID	Test case name	Test case	Output
1	Uploading relevant tweet	User uploads a tweet related to Cryptocurrency	Upload successful
2	Uploading relevant Reddit posts	User uploads Reddit posts related to Cryptocurrency	Upload successful
3	Uploading relevant News Headlines	User uploads News headlines related to Cryptocurrency	Upload successful

Table 6.1 Test Case Table for Training Data

6.3.2 TEST CASE 2:

Test case for Testing data:

The trained models are tested and their results are considered to fine tune the parameters of the model.

Test case ID	Test case name	Test case	Output
1	ETH	Price prediction	predicted price successfully
2	LTC	Price prediction	predicted price successfully
3	DOGE	Price prediction	predicted price successfully
4	BTC	Price prediction	predicted price successfully

Table 6.2 Test Case Table for Testing Data

7. CONCLUSION

7. CONCLUSION & FUTURE SCOPE

7.1 PROJECT CONCLUSION

Bitcoin Prediction is actual demand for beneficial business as the use of cryptocurrency is growing every day. Predictions are always helpful to decrease risk factors in any business environment. Risk factors can be analyzed on the basis of historical data and previous business trends. The results of this project will help us predict the trends beforehand when investing in any cryptocurrency and help us reduce the chances of facing a loss.

7.2 FUTURE SCOPE

Bitcoin's future is still very uncertain, but it's on the cusp of mainstream acceptance. The institutional investor interest is driving broad interest in cryptocurrency, but issues over custody, security, and capital efficiency are still headwinds for digital assets. With the help of this prediction model, investors will be able to know the changes in the prices of Bitcoin beforehand. Deep learning techniques such as the LSTM model combined with Sentiment analysis and prediction using Machine Learning, will help them reduce the chances of facing a loss. This system with little modification can be used to predict the price of any cryptocurrency in the near future.

With regards to future work, we seek to expand our dataset, both in terms of numbers and diversity, as a first step to achieving better results. More relevant labels can be added according to the status quo, and various deep learning models can be experimented on.

8. BIBLIOGRAPHY

8. BIBLIOGRAPHY

8.1 REFERENCES

[1] Hamayel, Mohammad J., and Amani Y. Owda. 2021. "A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and bi-LSTM Machine Learning Algorithms" AI 2, no. 4: 477-496. https://doi.org/10.3390/ai2040030

[2] Mallqui, Dennys C.A., and Ricardo A.S. Fernandes. 2018. "Predicting The Direction, Maximum, Minimum And Closing Prices Of Daily Bitcoin Exchange Rate Using Machine Learning Techniques". Applied Soft Computing 75: 596-606. doi:10.1016/j.asoc.2018.11.038.

[3] McNally, Sean, Jason Roche, and Simon Caton. 2018. "Predicting The Price Of Bitcoin Using Machine Learning". In 2018 26Th Euromicro International Conference On Parallel, Distributed And Network-Based Processing (PDP), 23. IEEE. https://ieeexplore.ieee.org/document/8374483.

[4] M, Vaidehi, Alivia Pandit, Bhaskar Jindal, Minu Kumari, and Rupali Singh. 2021. "BITCOIN PRICE PREDICTION USING MACHINE LEARNING". International Journal Of Engineering Technologies And Management Research 8 (5): 20-28. doi:10.29121/ijetmr.v8.i5.2021.953.

8.2 WEBSITES

1. https://stackoverflow.com/
2. https://scikit-learn.org/stable/
3. https://min-api.cryptocompare.com/
4. https://colab.research.google.com/

8.3 GITHUB REPOSITORY LINK

https://github.com/kiran-lenka/Bitcoin-Price-Predictor

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 04 | April 2022 www.irjet.net p-ISSN: 2395-0072

Bitcoin Price Prediction and Forecasting

K.Ranjith Reddy[1], Kiran Lenka[2], N. Devi[3] , D.V. Rithika[4]

[1]Assistant Professor, Computer Science and Engineering, JNTUH/CMRTC, Hyderabad, Telangana, India
[2,3,4]B.Tech Student, Computer Science and Engineering, JNTUH/CMRTC, Hyderabad, Telangana, India

---

Abstract - *Bitcoin is one of the oldest and biggest cryptocurrencies being traded as of now, in terms of the volume being traded. It is a digital asset over which central banks or any financial Institutions have no control or regulations. Bitcoin has a market share of more than 55% as compared to other cryptocurrencies. It is very sporadic and this is one of the most important reasons which attracted us to analyze and predict its price. Here, we intend to study the prediction of Bitcoin prices using Machine Learning Techniques and prepare a strategy to maximize gains for investors.*

Key Words: **Bitcoin prices, LSTM, Tweets, Reddit posts.**

1. INTRODUCTION

The stock market is one of the most volatile data available in terms of Machine learning datasets. Researchers have been long trying to predict the stock market and any breakthrough in this field would result in, literally, the people being able to mint money. Cryptocurrencies, to be specific, have gained a lot of traction in recent years from investors across the globe. There are several reasons why Bitcoin's price history has been so volatile. It is important to learn the factors that affect its market price so you can decide whether to invest, trade, or keep an eye on its development. As with most commodities, investments, assets, or other products, Bitcoin's price is heavily influenced by supply and demand. Since Bitcoin is a rapidly adopted asset by investors and traders, speculation about price movements plays a crucial role in its value at any given point in time.

Given the vastness of investments in Bitcoin, Cryptocurrencies are relatively unpredictable compared to traditional financial predictions like stock market predictions. Due to their high volatility, they offer a great opportunity for high profit if intelligent investing strategies are utilized. Here the Crypto-situation is analysed based on sentiment analysis from various platforms along with multiple learning approaches as a decent model will be able to provide the crypto investor with accurate data.

2. LITERATURE SURVEY

According to a research from 2016, more than 600 papers have been published on this topic. Our review of the literature focuses on work on bitcoin (BTC) price prediction utilising various methodologies, as well as the requirement for an evaluation of recurrent neural networks (RNN) and their system design. Dennys et al.[2] applied machine learning methods such as artificial neural networks (ANN), support vector machines (SVM), and recurrent neural networks (RNN), as well as K-means clustering, to get the most important features and used different attributes selection mechanisms to get the most important features. One limitation of this study is that it is solely focused on investors. Because bitcoin has the potential to transform the dynamics of the global economy, policymakers should be regarded as major partners in the system. Sean McNally et al. [3] used a Bayesian optimized recurrent neural network and LSTM to predict the direction of Bitcoin price in USD. They also compared the deep learning approaches using the ARIMA model.

Bitcoin purchases and sales can provide enormous amounts of benefits when done correctly. It has proven to be a fortune for many people in the past and is still making them a lot of money today. The downside of this is that it doesn't come without a price. Likewise, if not planned and calculated properly, you can lose a lot of money. You should have a great understanding of how and precisely why bitcoin price changes, which means you need to know how people make their bitcoin predictions. As well as these things (supply and demand, regulations, news, etc.), one must consider the technology of bitcoin and its progress. This aside, we now have to deal with the technical part using various algorithms and technologies which can predict precise bitcoin price movements. In terms of machine learning and deep neural networks, a time series is normally a sequence of numbers that are associated with a specific point in time. Since this is a time-series data set, it should be split into two parts: inputs and outputs. Further, LSTM can easily handle multiple input problems, compared to the classic statistics linear models. The idea of this project is to use Deep learning techniques such as the LSTM model combined with Sentiment analysis and prediction using Machine Learning techniques. Investors will be able to foresee changes in Bitcoin values with the support of this prediction model, which will also help them limit the risk of losing money. In order to analyze and predict sentiment, we will use a Machine Learning tool, which will analyze texts based on their polarity, ranging from positive to negative. With these examples of emotions in text, machines automatically learn how to detect sentiment without human input. Sentiment analysis on Reddit and Twitter are popular social media platforms where

discussion on CryptoCurrency is higher when compared to other sites. Analyzing data with visual elements such as charts, graphs, and maps and seeing patterns, outliers, and trends. Training and testing of dataset extracted from CryptoCompare tool to predict the price of Bitcoin using the LSTM neural network along with Sentiment analysis to analyze the model performance.

3. METHODOLOGY

To reach the goals of this study, we used historical cryptocurrency prices to train three separate models for three different types of cryptocurrency price prediction. Then, to assess the effectiveness of the proposed schemes, we compare the accuracy of our proposed model to that of existing models in five stages: (1) gathering historical AI 2021, 2 481 cryptocurrency data; (2) data exploration and visualisation; (3) training three different types of models; (4) testing the models; and (5) extracting and comparing the findings. In this section, we introduce and analyse three types of algorithms in this section: long short-term memory (LSTM), gated recurrent unit (GRU), and bidirectional LSTM (bi-LSTM)—to forecast the price of three types of cryptocurrency based on historical data: Bitcoin (BTC), Litecoin (LTC), and Ethereum (ETH) [1]. Sentiment analysis and market sentiment prediction on popular social media platforms such as Twitter and Reddit using tweets and posts related to Bitcoin will be performed. Bitcoin price data collected using the CryptoCompare API (provides highly reliable historical data for the public) and Data visualization will help investors to know the volatility in the prices of bitcoins.

Some of the previous studies' drawbacks are that they take into account a variety of features that aren't as useful in stock price prediction. Some studies miss many aspects that provide insight into the data. Existing efforts are best suited to less volatile stock market prices and cannot be used for extremely volatile cryptocurrencies like Bitcoin. Here we aim to train and test the dataset to predict the price of Bitcoin using the LSTM neural network. Visualize the prediction results in the form of a graphical representation that contains the information and the data. The proposed methodology has certain advantages because of which we will be able to predict the sentiment of tweets, news headlines and Reddit posts related to bitcoin, With the help of this system, one can predict the stock price of Bitcoin and make their investments carefully.

The risk of loss can be avoided by predicting the price of Bitcoin beforehand. This system with little modification can be used to predict the price of any cryptocurrency.

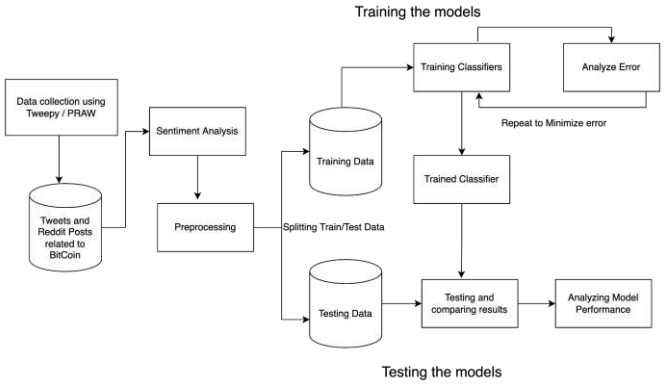


Figure 1: The model architecture of sentiment analysis

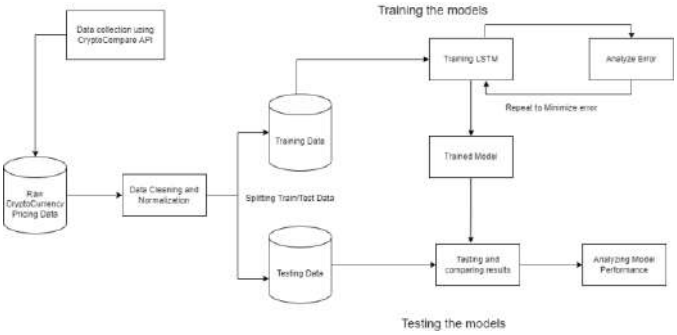


Figure 2: The model architecture of BTC price prediction

3.1 DESCRIPTION

Input Data: There are two forms of input data in this system. The first type of input data is the tweets of current trends and prices related to Bitcoin. It is a CSV file with approximately 8000 tweets. The second type of input data is the Reddit posts related to Bitcoin prices where content is socially curated and promoted by the social website members through upvoting. This is used to create, train, and assess the models' performance.

Data Cleaning and Preprocessing: In this step, the input data is cleaned and preprocessed. This step involves the removal of emoticons, hashtags, mentions, numerical values, punctuations and stop words, stemming and lemmatization.

Sentiment Analysis and Classification: Here the machine visualizes how many positive, neutral and negative tweets are present using bar plots, training and testing 6 different models. It then plots the performance between these models.

Data Visualization: In this stage, you'll see historical price movements, compare them to the Nifty-50 and the BSE-Sensex, see volatility, collect data with CryptoCompare, train and test the LSTM model, and see how it performs.

Building Models: Multiple machine learning algorithms are built based on the features extracted using various natural language processing techniques. Such as Logistic Regression, Random Forest, SGD, Support-Vector Machine, LightGBM, Gradient Boosting Classifier.

Evaluating Models: The models built are carefully evaluated on the test dataset using various performance metrics such as Precision, Accuracy, F1 Score etc. and are compared. The whole system can be divided into two major phases such as training and testing.

Training the models: Here multiple models are trained using the labelled dataset and this dataset is used to train Machine Learning Algorithms and is fed with sufficient training data to learn from.

Testing the models: The trained models are tested and their results are considered to fine-tune the parameters of the model. It is the process where the performance of a fully trained model is evaluated on a testing set.

3.2 LSTM IMPLEMENTATION

In [4], the researchers evaluated a key feature of feed networks is that they do not save memory. As a result, each input is processed independently, with no preserved state in the middle. We should maintain track of future events because we are dealing with a series of situations when information from the prior Bitcoin price is needed. Long short-term memory (LSTM) networks are a sort of recurrent neural network that can learn how to predict sequences based on their order. This is a necessary operation in complex problem areas such as machine translation and speech recognition.

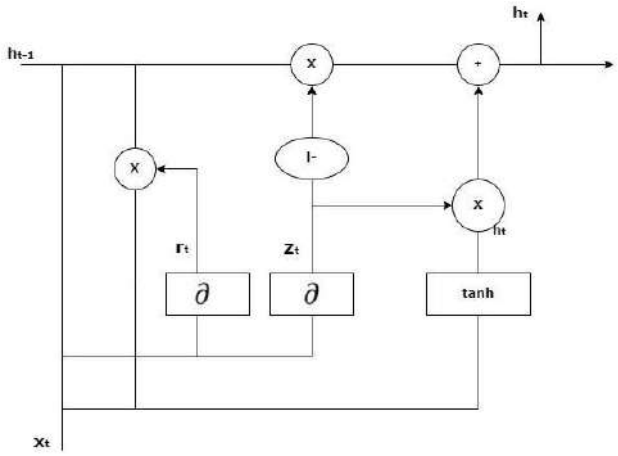


Figure 3: Long Short Term Memory Network

As LSTMs has grown in popularity, various changes to the traditional LSTM architecture have been attempted to simplify the internal design of the cell, make the cell work more efficiently, and reduce computational complexity. Gers and Schmidhuber have introduced a peephole connection where the gate layer can always know the state of the cell. Some LSTMs also used combined input and oblivion gates instead of two separate gates to help make both decisions at the same time. Another variation was the use of a gated regression unit (GRU). This has improved design complexity by reducing the number of gates. The combination of cell state and hidden state, and forgotten update gates and input gates are merged and used.

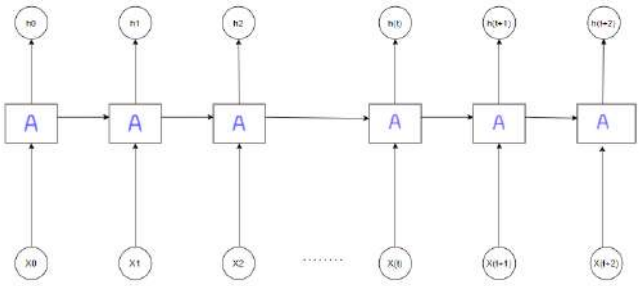
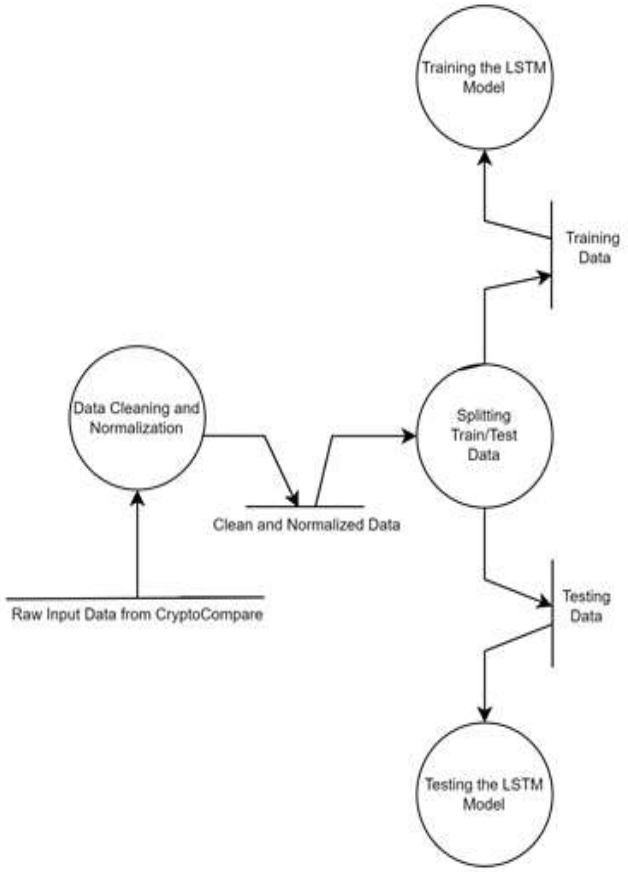


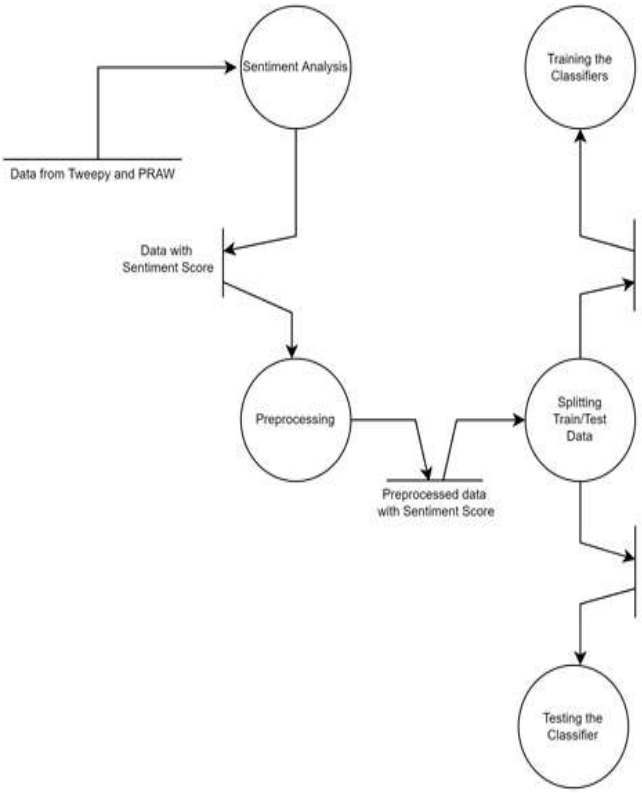
Figure 4: An unrolled recurrent neural network

4. SYSTEM DESIGN

4.1 DATA FLOW DIAGRAM DURING BTC PRICE PREDICTION USING LSTM

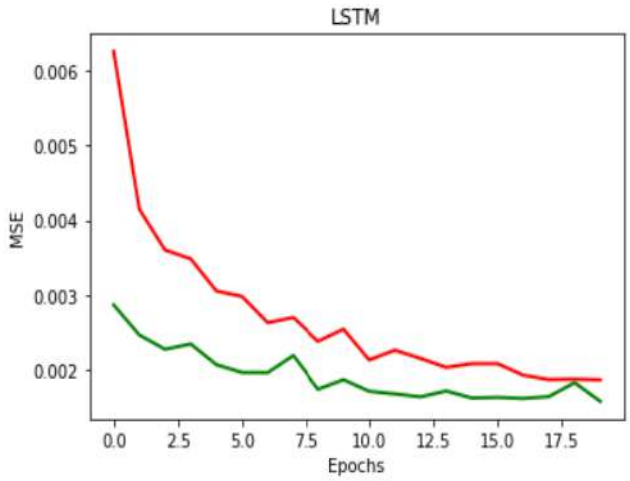


4.2 DATA FLOW DIAGRAM DURING BITCOIN MARKET SENTIMENT ANALYSIS USING DATA FROM TWITTER AND REDDIT POSTS

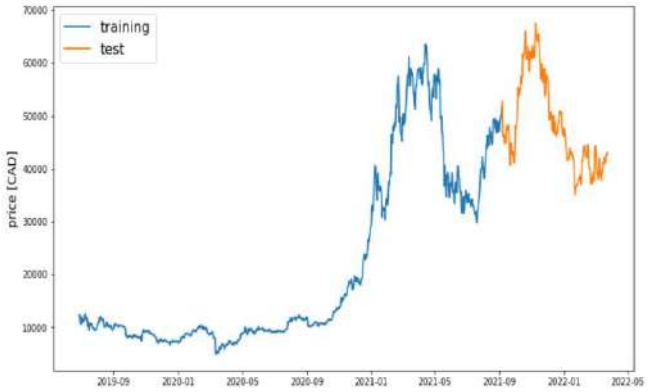


5. CRYPTOCURRENCY PRICES

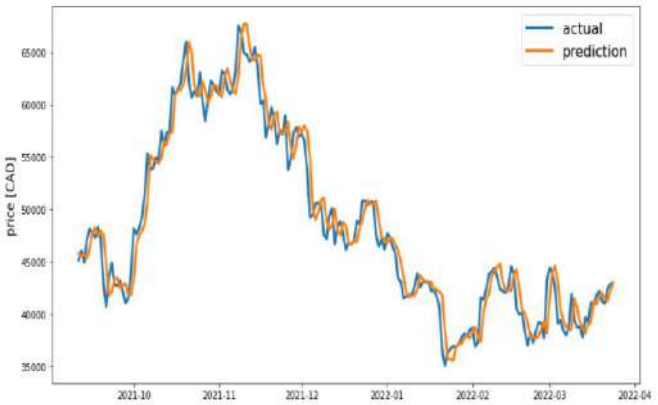
5.1 GRAPH REPRESENTING MSME AND EPOCHS USING LSTM



5.2 CRYPTOCURRENCY PRICES IN INDIAN RUPEES AS A FUNCTION OF TIME



5.3 GRAPH PLOT OF ACTUAL AND PREDICTED PRICES



6. CONCLUSION AND FUTURE SCOPE

BitCoin Prediction is the actual demand for beneficial business as the use of cryptocurrency is growing every day. Predictions are always helpful to decrease risk factors in any business environment. Risk factors can be analyzed on the basis of historical data and previous business trends. The results of this project will help us predict the trends beforehand when investing in any cryptocurrency and help us reduce the chances of facing a loss.

With the help of this prediction model, investors will be able to know the changes in the prices of Bitcoin beforehand. Deep learning techniques such as the LSTM model combined with Sentiment analysis and prediction using Machine Learning, will help them reduce the chances of facing a loss. This system with little modification can be used to predict the price of any cryptocurrency in the near future.

With regards to future work, we seek to expand our dataset, both in terms of numbers and diversity, as a first step to achieving better results. More relevant labels can be added according to the status quo, and various deep learning models can be experimented on.

REFERENCES

[1] Hamayel, Mohammad J., and Amani Y. Owda. 2021. "A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and bi-LSTM Machine Learning Algorithms" AI 2, no. 4: 477-496. https://doi.org/10.3390/ai2040030

[2] Mallqui, Dennys C.A., and Ricardo A.S. Fernandes. 2018. "Predicting The Direction, Maximum, Minimum And Closing Prices Of Daily Bitcoin Exchange Rate Using Machine Learning Techniques". Applied Soft Computing 75: 596-606. doi:10.1016/j.asoc.2018.11.038.

[3] McNally, Sean, Jason Roche, and Simon Caton. 2018. "Predicting The Price Of Bitcoin Using Machine Learning". In *2018 26Th Euromicro International Conference On Parallel, Distributed And Network-Based Processing (PDP)*, 23. IEEE. https://ieeexplore.ieee.org/document/8374483.

[4] M, Vaidehi, Alivia Pandit, Bhaskar Jindal, Minu Kumari, and Rupali Singh. 2021. "BITCOIN PRICE PREDICTION USING MACHINE LEARNING". International Journal Of Engineering Technologies And Management Research 8 (5): 20-28. doi:10.29121/ijetmr.v8.i5.2021.953.

e-ISSN: 2395-0056 p-ISSN: 2395-0072

International Research Journal of Engineering and Technology (IRJET)

(An ISO 9001 : 2008 Certified Journal)

Is hereby awarding this certificate to

Kiran Lenka

In recognition the publication of the manuscript entitled

Bitcoin Price Prediction and Forecasting

published in our Journal Volume 9 Issue 4 April 2022

Editor in Chief

E-mail : editor@irjet.net

Impact Factor : 7.529

www.irjet.net

e-ISSN: 2395-0056 p-ISSN: 2395-0072

International Research Journal of Engineering and Technology (IRJET)

(An ISO 9001 : 2008 Certified Journal)

Is hereby awarding this certificate to

N. Devi

In recognition the publication of the manuscript entitled

Bitcoin Price Prediction and Forecasting

published in our Journal Volume 9 Issue 4 April 2022

Editor in Chief

E-mail : editor@irjet.net

Impact Factor : 7.529

www.irjet.net

e-ISSN: 2395-0056 p-ISSN: 2395-0072

International Research Journal of Engineering and Technology (IRJET)

(An ISO 9001 : 2008 Certified Journal)

Is hereby awarding this certificate to

D.V. Rithika

In recognition the publication of the manuscript entitled

Bitcoin Price Prediction and Forecasting

published in our Journal Volume 9 Issue 4 April 2022

Editor in Chief

E-mail : editor@irjet.net

Impact Factor : 7.529

www.irjet.net