BITCOIN PRICE PREDICTION USING DEEP LEARNING AND SENTIMENT ANALYSIS

BY

VIBITHA.M

(19PCS019)

Project Report Submitted

In partial fulfillment of the requirements for the Award of

Master's Degree in Computer Science

DEPARTMENT OF COMPUTER SCIENCE

AVINASHILINGAM INSTITUTE FOR HOME SCIENCE AND

HIGHER EDUCATION FOR WOMEN (Deemed to be University)

COIMBATORE – 641043

MAY - 2021

BITCOIN PRICE PREDICTION USING DEEP LEARNING AND SENTIMENT ANALYSIS

BY

VIBITHA.M

(19PCS019)

Project Report Submitted

In partial fulfillment of the requirements for the Award of

Master's Degree in Computer Science

DEPARTMENT OF COMPUTER SCIENCE

AVINASHILINGAM INSTITUTE FOR HOME SCIENCE AND

HIGHER EDUCATION FOR WOMEN (Deemed to be University)

COIMBATORE – 641043

Signature of the
Head of the Department

Signature of Supervisor

Viva-voce Examination Held on __________________

Signature of Examiner

ACKNOWLEDGEMENT

I would like to express my sincere thanks to **God** Almighty, for his constant love and grace that he has showed upon me, which kept me in good health, and sound mind without which my project would not have reached a successful end.

I would like to express my deep sense of reverential gratitude and sincere thanks to **Dr. S. P. Thyagarajan ,** Chancellor, Avinashilingam Institute of Home Science and Higher Education for Women, Coimbatore, for the opportunity given to me for undertaking this study and for providing all the needed facilities during the course of my study.

I owe my great deal of gratitude to **Dr. Premavathy Vijayan, M.Sc., M.Ed., Dip.Spl.Edn., M.Phil., Ph.D**., Vice Chancellor, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for extending all resources that facilitated the conduct of the present study.

I express my gratitude to **Dr. S. Kowsalya, Registrar, M.Sc., M.Phil., Ph.D** Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for providing all facilities necessary for the study.

I would express my boundless thanks to **Dr. K. Udaya Chandrika, M.Sc., M.Phil., Ph.D., Dean**, School of Physical Sciences & Computational Sciences, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for granting the facility required.

I wish to place on record my deep sense of gratitude to **Dr. Vasantha Kalyani David, M.Sc., M.Phil(Maths)., M.Phil(CS)., Ph.D, Professor and Head,** Department of Computer Science for support and encouragement to complete the project.

I express my heart full gratitude to my esteemed mentor **Dr. I. Elizabeth Shanthi, M.Sc., M.Phil., Ph.D., Professor, Department of Computer Science,** for imparting the tremendous assistance and well-timed support for triumph of our project with guidance and constant supervision as well as for providing necessary resources for the project and also for her support in completing the project.

I am grateful to the project coordinator **Dr. G. Sudhamathy, M.C.A., Ph.D., Assistant Professor**, Department of Computer Science, who was instrumental in granting me the facilities required for doing project.

Finally, yet importantly, I would like to thank my **parents, family members and friends** for their kind inspiration, support, encouragement, blessings and prayers, which were instrumental in the successful completion of the project.

I have great pleasure in expressing my deep sense of gratitude to all other teaching and non-teaching staff members of the Department of Computer Science, who stood behind the screen for the completion of the project.

I would extend my hearty thanks to one and all that helped me directly or indirectly for the successful completion of my project.

ABSTRACT

Bitcoin is the first digital decentralized cryptocurrency that has shown a significant increase in market capitalization in recent years. The objective of this project is to determine the predictable price direction of Bitcoin in USD by deep learning techniques and sentiment analysis. Twitter have attracted a great deal of attention from researchers to study public sentiment. We have applied sentiment analysis and supervised machine learning principles to the extracted tweets from Twitter and we analyze the correlation between bitcoin price movements and sentiments in tweets. We explored algorithm of deep learning to develop a prediction model and provide informative analysis of future market prices. Due to the difficulty of evaluating the exact nature of a Time Series(ARIMA) model, it is often very difficult to produce appropriate forecasts. Then we continue to implement Recurrent Neural Networks (RNN) with long short-term memory cells (LSTM). Thus, we analyzed the time series model prediction of bitcoin prices with greater efficiency using long short-term memory (LSTM) techniques and compared the predictability of bitcoin price and sentiment analysis of bitcoin tweets to the standard method (ARIMA).

CONTENTS

INTRODUCTION

INTRODUCTION

1.1PROBLEM DEFINITION

Bitcoin price prediction has been an active area of research for a long time. Bitcoin, as a pioneer within the block chain monetary renaissance plays an overwhelming part in an entirety cryptocurrency market capitalization environment. Hence, it is the incredible interest of deep learning and data mining community to be able to:

(I) Predict Bitcoin price changes

(II) Grant experiences to get it what drives the Bitcoin instability and way better assess related dangers in cryptocurrency domain. Many researchers worked on machine learning algorithms and sentiment analysis from social media to find out the bitcoin stock market price prediction.

(III) In our project, we used a time series model and deep learning to leverage machine learning technology to predict the real-time price of Bitcoin. However, machine learning literature is lacking verification of whether or not the stock evaluation strategies are legitimate for the cryptocurrencies, and if so, how they may be modified.

(IV) That is what features want to be eliminated or introduced as a foundation for price prediction, whether current machine learning algorithms work for cryptocurrencies, and which technique yields the excellent outcomes. In this paper, we check out these questions.

(V) Such evaluation is applicable given an extremely good quantity of attention that cryptocurrencies, particularly Bitcoin, are generating. Each people and big financial corporations are drawn to cryptocurrencies due to the transparency and anonymity that they offer to their users, in addition to their resistance to fraud because of the dispersed nature of the ledger statistics.

(VI) Moreover, purchasing cryptocurrencies are promising in terms of making income, and ought to be of interest to buyers. In addition, to familiarizing themselves with industry tendencies and political and financial information, they can utilize machine learning models to decide whether to buy or promote cryptocurrency

1.2 PROBLEM STATEMENT

1. Bitcoin is the most complex cryptocurrency which value change in every second.

2. Investing money for Bitcoin is more risk and less profit.

1.3 PROJECT OBJECTIVES

1. To predict bitcoin price with maximum efficiency using LSTM and ARIMA.

2. To ensure less risk and more profit for investors.

1.4 OVERVIEW OF THE PROJECT

In this project, we used a time series model and deep learning technology to predict the real-time price of Bitcoin. However, deep learning literature is lacking verification of whether or not the stock evaluation strategies are legitimate for the cryptocurrencies, and if so, how they may be modified. That is what features want to be eliminated or introduced as a foundation for price prediction, whether current machine learning algorithms work for cryptocurrencies, and which technique yields the excellent outcomes. Such evaluation is applicable given an extremely good quantity of attention that cryptocurrencies, particularly Bitcoin, are generating. Each people and big financial corporations are drawn to cryptocurrencies due to the transparency and anonymity that they offer to their users, in addition to their resistance to fraud because of the dispersed nature of the ledger statistics. Moreover, purchasing cryptocurrencies are promising in terms of making income, and ought to be of interest to buyers. In addition, to familiarizing themselves with industry tendencies and political and financial information, they can utilize deep learning models to decide whether to buy or promote cryptocurrency.

SYSTEM SPECIFICATION

2. SYSTEM SPECIFICATON

To develop “BITCOIN PRICE PREDICTION USING DEEP LEARNING AND SENTIMENT ANALYSIS” the following hardware and software configuration is used:

2.1 HARDWARE SPECIFICATION

Processor: AMD PRO A4-3350B APU with Radeon R4 Graphics 2.00GHz

Installed Memory (RAM): 4GB

System Type: 64-bit Operating System, x64 based processor

2.2 SOFTWARE SPECIFICATION

Operating System: Windows 10 Pro

Software: Python 3.7 – Anaconda3-5.3.1- Windows-X86_64

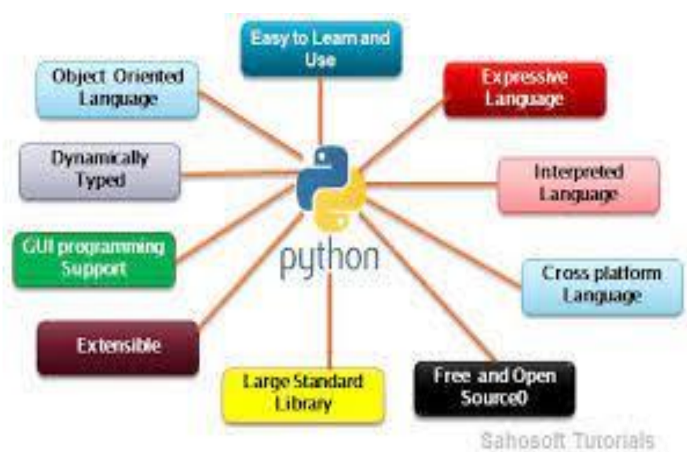
2.3 ABOUT THE SOFTWARE

2.3.1 PYTHON

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace. It provides constructs that enable clear programming on both s mall and large scales. Van Rossum led the language community until stepping down as leader in July 2018. Python can be used on a server to create web applications.

Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object oriented, imperative, functional and procedural, it also has a comprehensive standard library. Python interpreters are available for many operating systems. CPython, the reference implementation of Python, is open source software and has a community-based development model, as do nearly all of Python's other implementations. Python and CPython are managed by the non-profit Python Software Foundation.

Python is a multi-paradigm programming language. Object-oriented programming and structured programming are fully supported, and many of its features support functional programming and aspect-oriented programming (including by meta programming and meta objects (magic methods)). Many other paradigms are supported via extensions, including design by contract and logic programming.



2.3.2 PYTHON PACKAGES

1) Numpy

Numpy will help you to manage multi-dimensional arrays very efficiently. Maybe you won't do that directly, but since the concept is a crucial part of data science, many other libraries (well, almost all of them) are built on Numpy. Simply put: without Numpy you won't be able to use Pandas, Matplotlib, Scipy or Scikit-Learn. That's why you need it on the first hand.

```
In [1]: import numpy as np

In [2]: a = np.arange(12).reshape(2, 2, 3)

In [3]: a
Out[3]: array([[[ 0,  1,  2],
                [ 3,  4,  5]],

               [[ 6,  7,  8],
                [ 9, 10, 11]]])
```

2) Pandas

To analyze data, we like to use two-dimensional tables – like in SQL and in Excel. Originally, Python didn't have this feature. Weird, isn't it? But that's why Pandas is so important! I like to say, Pandas is the "*SQL of Python.*" (Eh, I can't wait to see what I will get for this sentence in the comment section… ;-)) Okay, to be more precise: Pandas is the library that will help us to handle two-dimensional data tables in Python. In many senses it's really similar to SQL, though.

```
In [6]: super_tree.head()
Out[6]:
```

	day	my_date	user_id	event_type
0	day_1	2017-12-01	1000007	sent_a_super_tree
1	day_1	2017-12-01	1000010	sent_a_super_tree
2	day_1	2017-12-01	1000011	sent_a_super_tree
3	day_1	2017-12-01	1000019	sent_a_super_tree
4	day_1	2017-12-01	1000022	sent_a_super_tree

3) Matplotlib

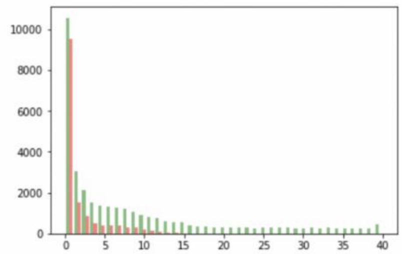
I hope I don't have to detail why data visualization is important. Data visualization helps you to better understand your data, discover things that you wouldn't discover in raw format and communicate your findings more efficiently to others.

The best and most well-known Python data visualization library is Matplotlib. I wouldn't say it's easy to use… But usually if you save for yourself the 4 or 5 most commonly used code blocks for basic line charts and scatter plots, you can create your charts pretty fast.

VISUALIZATION

```
In [14]: android = big_table[big_table.phone_type == 'android'].reset_index()
         ios = big_table[big_table.phone_type == 'ios'].reset_index()
```

```
In [15]: bins = np.linspace(0, 40, 40)
         x = android['free']
         y = ios['free']
         data = [x,y]
         plt.hist(data, bins, alpha = 0.5, color = ['g','r'])
         plt.show()
```



4) Scikit-Learn

Without any doubt the fanciest things in Python are Machine Learning and Predictive Analytics. And the best library for that is Scikit-Learn, which simply defines itself as "Machine Learning in Python." Scikit-Learn has several methods, basically covering everything you might need in the first few years of your data career: regression methods, classification methods, and clustering, as well as model validation and model selection. You can also use it for dimensionality reduction and feature extraction.

(Get started with my machine learning tutorials here: Linear Regression in Python using sklearn and numpy!)

5) Scipy

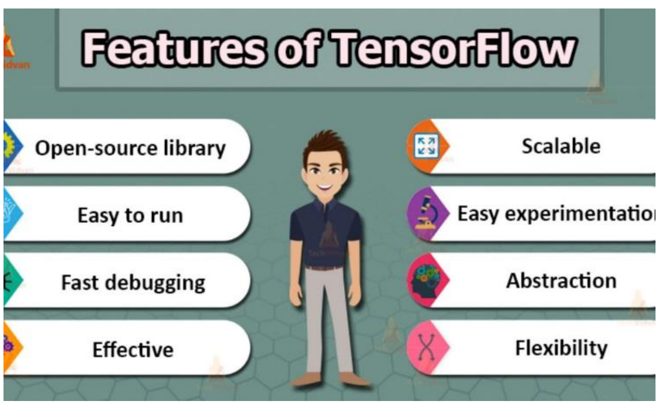
This is kind of confusing, but there is a Scipy *library* and there is a Scipy *stack*. Most of the libraries and packages I wrote about in this article are part of the Scipy *stack* (that is for scientific computing in Python). And one of these components is the Scipy *library* itself, which provides efficient solutions for numerical routines (the math stuff behind machine learning models). These are: integration, interpolation, optimization, etc.

Just like Numpy, you most probably won't use Scipy itself, but the above-mentioned Scikit-Learn library highly relies on it. Scipy provides the core mathematical methods to do the complex machine learning processes in Scikit-learn. That's why you have to know it.

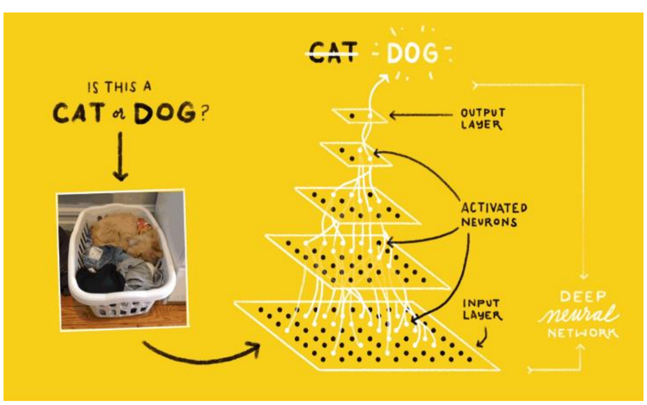
6) Tensorfow

TensorFlow is a Python library for fast numerical computing created and released by Google.It is a foundation library that can be used to create Deep Learning models directly or by using wrapper libraries that simplify the process built on top of TensorFlow.In this post you will discover the TensorFlow library for Deep Learning.

The main features of Tensor flow



This is how Tensorflow works



7) Tweepy

Tweepy is an open source Python package that gives you a very convenient way to access the Twitter API with Python. Tweepy includes a set of classes and methods that represent Twitter's models and API endpoints, and it transparently handles various implementation details, such as: Data encoding and decoding.

8) Textblob

TextBlob is a Python (2 and 3) library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more

9) Seaborn

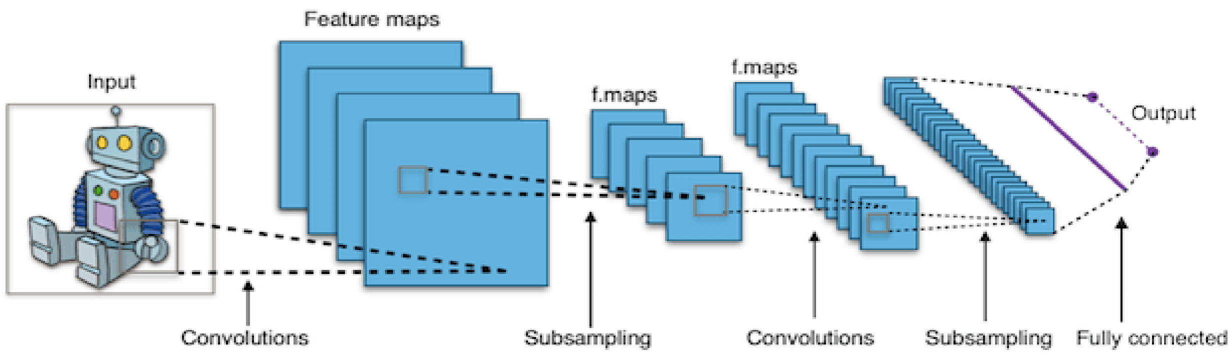
Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

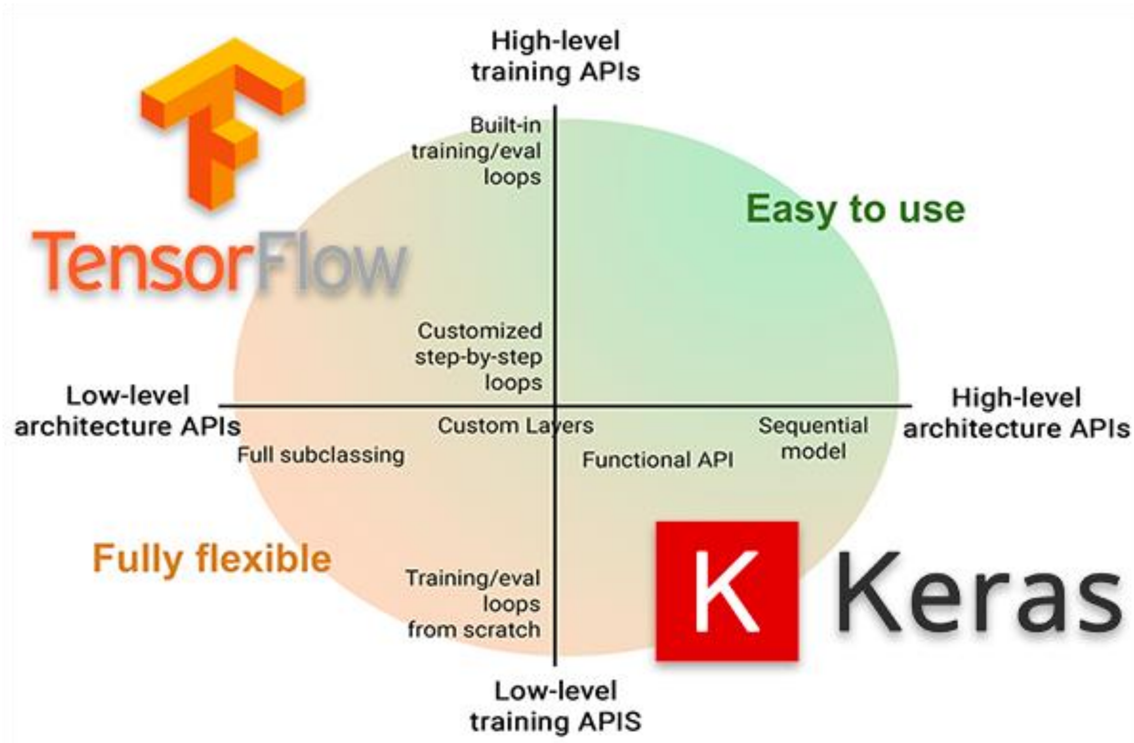
10) Keras

Keras is an Open Source Neural Network library written in Python that runs on top of Theano or Tensorflow. It is designed to be modular, fast and easy to use. It was developed by François Chollet, a Google engineer. Keras doesn't handle low-level computation.

These are the pros and cons of keras.

This is how keras works


Feature maps
Input
f.maps
f.maps
Output
Convolutions
Subsampling
Convolutions
Subsampling
Fully connected


High-level
training APIs
Built-in
training/eval
loops
TensorFlow
Easy to use
Customized
step-by-step
loops
Low-level
architecture APIs
Custom Layers
Full subclassing
Functional API
Sequential
model
High-level
architecture APIs
Fully flexible
Training/eval
loops
from scratch
Keras
Low-level
training APIS

SYSTEM DEVELOPMENT

3. SYSTEM DEVELOPMENT

3.1 DATASET DESCRIPTION

For the first phase of this paper, we have collected Bitcoin values from databases: The the real bitcoin dataset (Table 3.1) have taken from coinbase website (https://www.coinbase.com/price/bitcoin) .This dataset was gathered an interval length of every one-day and it continued the data collection process. From Coinbase API(1226824226109214723-iDcietRKZsDfGiIJQ9AiKkexlDml4U) we collected bitcoin real-time twitter data (Table 3.2) in order to predict the fluctuation of the bitcoin price with others collected dataset. We collected bitcoin price data by building an automated web scraper which is real-time twitter data pulled from Coinbase API over the course of weeks or months or years depending on program continuation. All these key features are connected with bitcoin network and important to real-time observation of moving prices.

BITCOIN DATASET FEATURES

Table 3.1

FEATURES	DATABASES	DESCRIPTIONS
Currency	Coinbase	Bitcoin
Date	Coinbase	Currenct date on that one minute time
Closing price(USD)	Coinbase	Closing price of the bitcoin that particular date
24^{th} Open(USD)	Coinbase	The open price of tge bitcoin on that one-minute time
24^{th} High(USD)	Coinbase	The highest price of bitcoin on that one-minute time
24^{th} Low(USD)	Coinbase	The lowest price of the bitcoin on that one-minute time

TWEET DATASET FEATURES

Table 3.2

FEATURES	DATABASE	DESCRIPTIONS
Tweet	Coinbase API	Bitcoin twitter tweet on that particular date
Date	Coinbase API	Current date of the tweet

3.2 MODULE DESCRIPTION

3.2.1 BITCOIN DATASET PROCESSING

1. **Cryptocurrency data overview**
2. **Time Series**
3. **Data preprocessing**
4. **Build and train LSTM model in Tensor Flow**
5. **Use the model to predict future Bitcoin price**

For the first phase of this project, we have collected Bitcoin values from four databases:
We gathered Bitcoin historical and real-time price data using their publicly available API. This dataset was gathered an interval length of every one-minute and it continued the data collection process. From Coinbase API we collected bitcoin real-time data in order to predict the fluctuation of the bitcoin price with collected dataset. We collected bitcoin price data by
building an automated web scraper which is real-time data pulled from Coinbase API over the course of weeks or months or years depending on program continuation. All these key features are connected with bitcoin network and important to real-time observation of moving prices.Figure 3.3

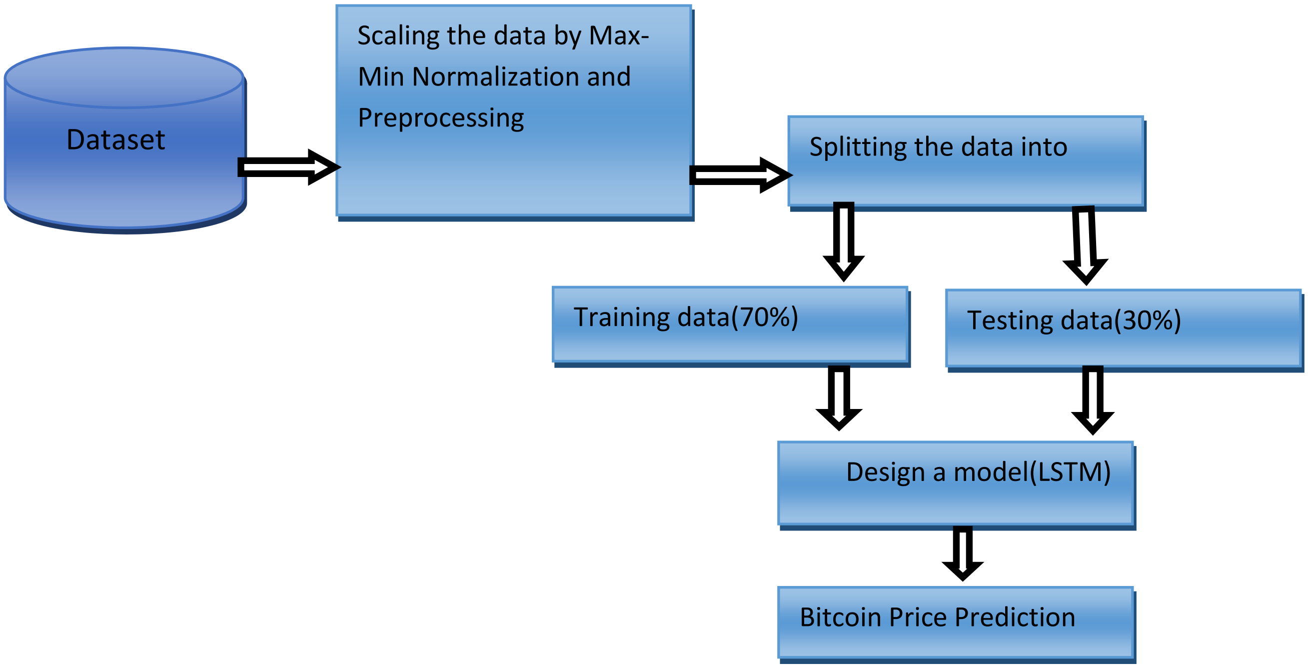


Figure 3.3 Bitcoin Prediction Flowchart

METHODOLOGY

Implementation

The study focuses primarily on the Bitcoin closing price for the development of the predictive model. The increase or decrease in the Bitcoin price with the higher volatility makes it harder to predict, but the deep learning models try to predict with some degree of accuracy. The implementation here is performed using the LSTM Recurring Neural Network.

Data Loading & Preparation

The data is loaded from the .csv file using the panda library function read_csv()(Figure 2 from appendix). The missing values are identified, and it is prepared for modeling by removing the unused fields and filling the NAN values.

Data Normalization

We need to normalize inputs in neural networks and other data mining models, otherwise, the network will perform poorly. Normalization is carried out in order to have the same range of values for each RNN model input. This can guarantee stable weight and partiality convergence. Normalization here uses the MinMaxScalar Package, after normalization, data is plotted using mat plot libraries and the trend is seen to check the fluctuations in the price and volume of the bitcoin over the last 1 years (2020-2021).

Data Splitting

We divide the data into training and test data in statistics and machine learning. The model is intended to fit the training data to make predictions. When we do this, there are chances of two things happening, one overfitting the model and the other under fitting the model. Overfitting means that the model is too well trained, and the predictions are too close and that it does not fit closely with the model when the model is under fitting. The Scikit Library is used here to split the data. The data for training is divided into 70 percent and 30 percent for testing. Total observation in the training dataset is 275 and 90 in the test dataset.

Bitcoin Prediction

After splitting the data validation loss (Figure 5) and prediction data (Figure 6). As we see Prediction Plot will decrease the price of bitcoin in future.

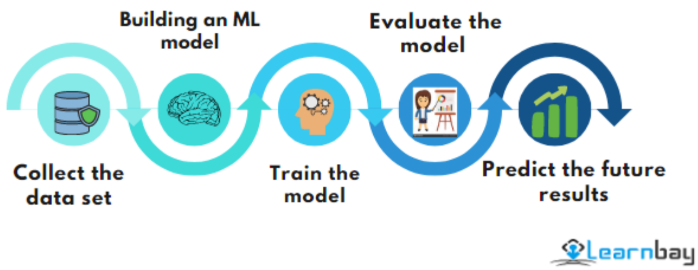


Figure 3.4 Process

3.2.2 SENTIMENT PROCESSING

Tweets Preprocessing

Tweets consist of many acronyms, emoticons and unnecessary records like images and URL's. So, tweets(Figure 3 From Appendix) are preprocessed to symbolize accurate feelings of the public. For preprocessing of tweets, we employed 3 ranges of filtering: Tokenization, Stopwords elimination, and Regex matching for removing special characters.

1. **Tokenization**: Tweets are split into character words primarily based on the gap and irrelevant symbols like emoticons are removed. We shape a list of individual words for each tweet(Figure 8).

2. **Stopword elimination**: Words that don't explicit any emotion is called Stopwords. After splitting a tweet, words like a, is, the, with etc., are eliminated from the listing of phrases.

3. **Regex matching for removing special characters**: Regex matching in Python is completed to suit URLs and are changed through the time period URL. Regularly tweets consist of hashtags (#) and @ addressing other users. They may be additionally changed definitely. For instance, #Microsoft is replaced with Microsoft and

@Billgates is changed with User. Extended word showing extreme feelings like coooooooool! is changed with cool!After those tiers, the tweets are ready for sentiment classification (Figure 9).

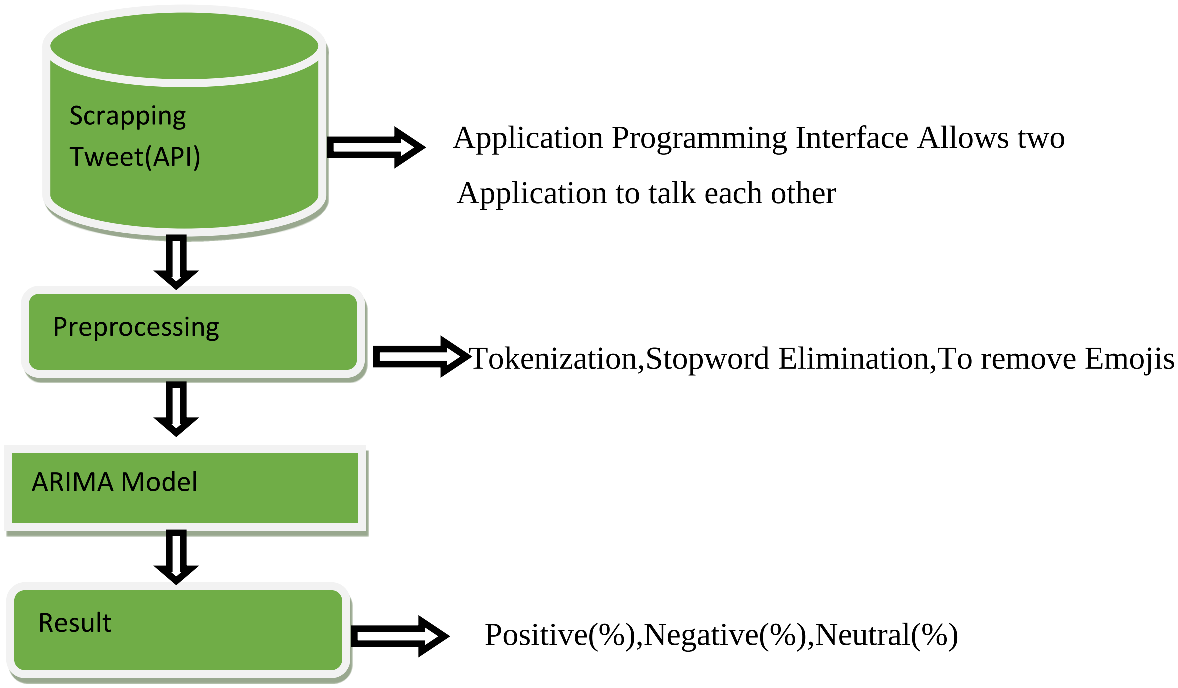


Figure 3.5 Tweet Processing Flowchart

Tweets Sentiment Analysis

Tweets are classified based on the sentiment as Positive (polarity >0), Negative (polarity <0) and Neutral (polarity =0). For individually tweet sentiment score, we used Textblob to automatically be passed the tweet text for analysis sentiment and gives polarity score. In this project we got following polarities that will shown in (Figure 10)

Zero polarities are 389

Positive polarities are 310

Negative polarities are 101

Number of subjective tweets : 162

Average sentiment is: 0.09396945120851367

3.2.3 MERGE THE DATASET

The Merging of Dataset

In this phase, we combine two different dataset (Figure 14) into single dataset by using merge data program that uses Pandas library. Merged dataset had only price, sentiment, and time features in our case. After merging the data we will be finding correlation. Which is Co-variance correlation, pearson correlation and spearman correlation to find accuracy. Figure 3.6

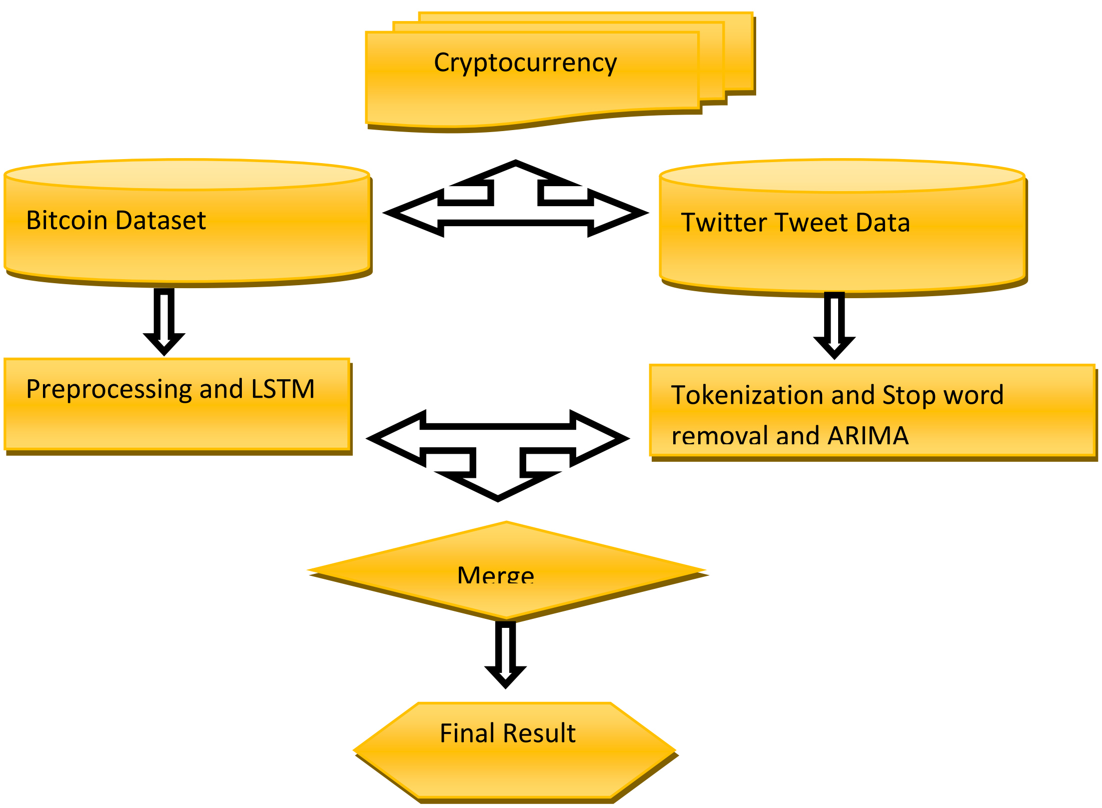


Figure 3.6 Merging the two dataset

3.2.4 FEATURE VECTOR

LSTM

The long short-term memory network or LSTM addresses the common problem of disappearing gradients in the recurrent neural network. This is a type of recurrent neural network that is used in profound learning, as very large architectures can be trained. LSTM enables the network to learn more about many time steps by maintaining a more-steady error. This enables the network to learn long-term trust. LSTM cell contains forget and remember gates that allow the cell to decide which information to block or transmit based on its strength and importance . As a result, weak signals that prevent the gradient from disappearing can be blocked. The performance of the RNN with LSTM network is assessed to determine the model's efficiency.

Elman's recent development of recurrent neural networks has gained popularity in network designs and increased computational power from graphical processing units. They are primarily useful with sequential data (in our case Bitcoins time series data (Figure 4)) because each neuron or unit can access its internal memory to keep information about the previous input. Figure 3.7 shows simple RNN Structure. One limitation of RNN is that it is influenced by the disappearing problem of the

gradient. This problem is that since the layers and time steps of the network are interrelated, they are susceptible to exploding or disappearing gradients.

$$f_t = \sigma_g(W_f * x_t + U_f * h_{t-1} + V_f \circ c_{t-1} + b_f)$$
$$i_t = \sigma_g(W_i * x_t + U_i * h_{t-1} + V_i \circ c_{t-1} + b_i)$$
$$c_t = f_t \circ c_{t-1} + i_t \circ \sigma_c(W_c * x_t + U_c * h_{t-1} + b_c)$$
$$o_t = \sigma_g(W_o * x_t + U_o * h_{t-1} + V_o \circ c_t + b_o)$$
$$h_t = o_t \circ \sigma_h(c_t)$$

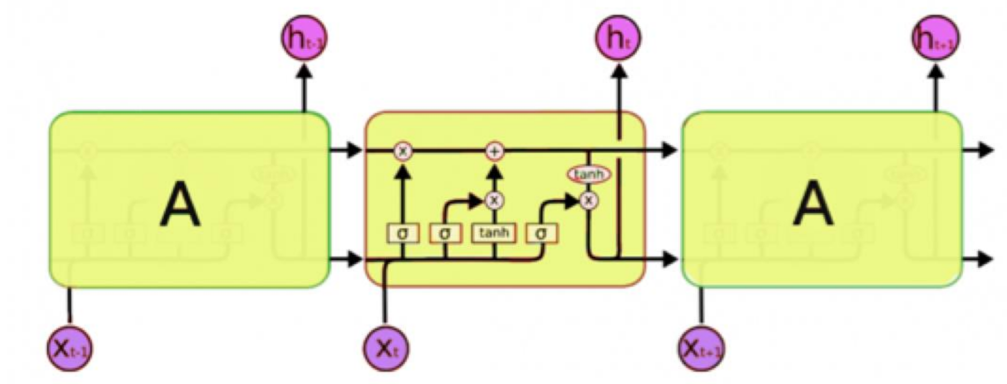


Figure 3.7

Vanishing gradients are a problem because they can prove to be too little to learn for the system, while inclinations can be restricted by regularization. In addition, some examinations have found that while RNN is able to deal with long-term dependencies, they often fail to learn in practice due to difficulties between gradient succession and long-term dependency.

ARIMA

The basic principle of the ARIMA model is to estimate the trend and the seasonality of the series and to remove them from the series (Figure 3.8) in order to obtain a stationary series. In this series, statistical forecasting techniques can be used. The final step would be to convert the forecast values to the original scale by applying constraints on trend and seasonality.

Trend – mean varying over time. For example, in this case, we saw an average decrease in the number of bitcoin price over time (Figure 12). Seasonality – time frame variations. For example, people may tend to buy cars in a given month due to pay increments or festivals.

ARIMA (Auto regression integrated moving average) is a class model that captures a suite of different standard temporal structures in time series data which include trend, seasonality, cycles, errors and nonstationary data. This allows it to exhibit dynamic temporal behavioral a time sequence.

MEAN SQUARED ERROR

Time serious usually focus on predicting real values, which are called regression problems. Hence, The performance measurements in focus on real-value prediction evaluation methods.
The most commonly used mean squared error.

Test Mean Squared Error of ARIMA (Figure 13) is 0.173.

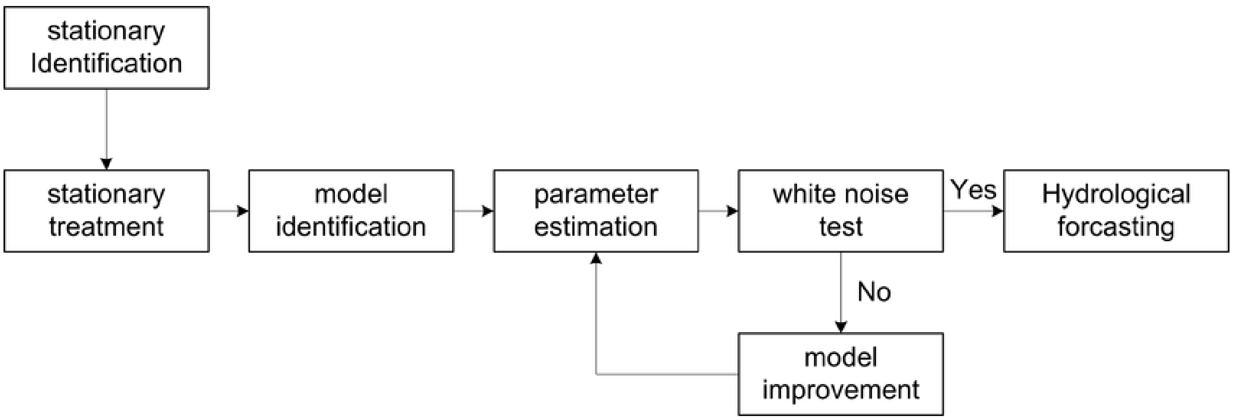


Figure 3.8

RESULTS & DISCUSSION

5. RESULT AND DISCUSSION

In order to decide proper strategy of analysis for studying the connection amongst Bitcoin's price and others significant parameters as well as sentiment analysis, the available associated literature co-variance matrix(Figure 15)- In this case where covariance matrix is positive, we can expect that the variables change in the same direction while if negative, the change will be in opposite direction. Karl Pearson's coefficient for correlation(Figure 16)

As we can see, there is a maximum of 0.14 correlation(-ve) in the 24h Low (USD) which is significantly low. Therefore we can conclude that these two variables are in low correlation with each other. calculate Spearman's correlation coefficient(Figure 17)- As we can see, there is less correlation (negative) between the pairs of data. The spearman's correlation ranges from -1 to +1. From above all the calculations gave negative results, so Bitcoin future price will be decrease.

CONCLUSION

6. CONCLUSION

This study focuses on the Bitcoin closing price and sentiments of the current market for the development of the predictive model. It does also calculate the market sentiments to predict the price more accurately. The prediction is limited to previous data. The ability to predict data streaming would improve the model's performance and predictability. The model developed using LSTM is more accurate than the traditional model. In our case, LSTM (Long Short-Term Memory) is obviously an effective learner on training data than ARIMA, with the LSTM more capable of recognizing long-term dependencies. The below table (table 6.1) most of the values negative correlation. The inference of the analysis is that the bitcoin price will decrease in the future.

Table 6.1 Correlation Calculation

Correlation	**24h High (USD)**	**24h Low (USD)**	**24h Open (USD)**	**Closing Price(USD)**
Co-Variance Correlation	6.64013363e+06, -1.57997393e+01, -1.57997393e+01, 1.23545306e-02	9.55651047e+06, -9.01058902e+01, -9.01058902e+01, 1.23545306e-02	7.53538098e+06, -4.91304367e+01, -4.91304367e+01, 1.23545306e-02	6.98404943e+06, 4.19822517e+01, 4.19822517e+01, 1.23545306e-02
Pearson Correlation	-0.05516306472	-0.2622348372	-0.1610218797	0.14292195
Spearman Correlation	0.0	-0.309523809	-0.1190476190	0.380952380

SCOPE FOR FUTURE ENHANCEMENT

7. SCOPE FOR FUTURE ENHANCEMENT

This study focuses on the Bitcoin closing price and sentiments of the current market for the development of the predictive model. It does also calculate the market sentiments to predict the price more accurately. The prediction is limited to previous data. The ability to predict data streaming would improve the model's performance and predictability. The model developed using LSTM is more accurate than the traditional models that demonstrate a deep learning model. In our case, LSTM (Long Short-Term Memory) is obviously an effective learner on training data than ARIMA, with the LSTM more capable of recognizing long-term dependencies. This study uses the daily price fluctuations of the Bitcoin to further investigate the model's predictability with hourly price fluctuations in the future. This paper consists only of comparing ARIMA with LSTM. The result would be confirmed by comparing more machine learning models in the future.

In this work, we have only considered Twitter posts data to analyze people's feelings that may be biased because not all people who trade in stocks share their views on Twitter. Moreover, Facebook posts and LinkedIn data can be included in a comprehensive collection of public opinion. In addition, the current sentiments can be combined with the prediction of the LSTM model to influence the decision of an autonomous trading assistant to buy or sell Bitcoins.

BIBLIOGRAPHY

8. BIBLIOGRAPHY

- https://www.researchgate.net/publication/342464470_Real-Time_Prediction_of_BITCOIN_Price_using_Machine_Learning_Techniques_and_Public_Sentiment_Analysis
- https://www.researchgate.net/publication/345262720_Bitcoin_Price_Prediction_and_Analysis_Using_Deep_Learning_Models
- https://github.com/rohanasthana/Bitcoin-Sentiment-Analysis/blob/master/Bitcoin%20Analysis.ipynb
- https://ieeexplore.ieee.org/abstract/document/9206704
- https://www.coinbase.com/price/bitcoin
- Phaladisailoed T, Numnonda T (2018) Machine learning models comparison for bitcoin price prediction. In: 2018 10th international conference on information technology and electrical engineering (ICITEE). IEEE, pp 506–511
- Madan I, Saluja S, Zhao A(2015) Automated bitcoin trading via machine learning algorithms. 1–5.
- http://cs229.stanford.edu/proj2014/Isaac\% 9. Lahmiri S, Bekiros S (2019) Cryptocurrency forecasting with deep learning chaotic neural networks. Elsevier, 35–40 10.
- SaxenaA,SukumarTR(2018)PredictingbitcoinPriceusinglstmandcompareitspredictability with Arima model. Int J Pure Appl Math 2591–2600 11. Paresh Kumar N, Narayan S, Rahman RE, Setiawan I (2019) Bitcoin price growth and Indonesia's monetary system. Emerg Mark Rev 38:364–376 12.
- PantDR,NeupaneP,PoudelA,PokhrelAK,LamaBK(2018)Recurrentneuralnetworkbased bitcoin price prediction by twitter sentiment analysis.
- IEEE, pp 128–132 13. Nivethitha P, Raharitha P (2019) Future stock price prediction using LSTM machine learning algorithm. Int Res J Eng Technol (IRJET) 1182–1186 14.
- Roth N (2015) An architectural assessment of bitcoin: using the systems modeling language. Procedia Comput Sci 44:527–536

APPENDIX

APPENDIX

A) WORKFLOW MODEL- This is the whole workflow diagram (Figure 1).

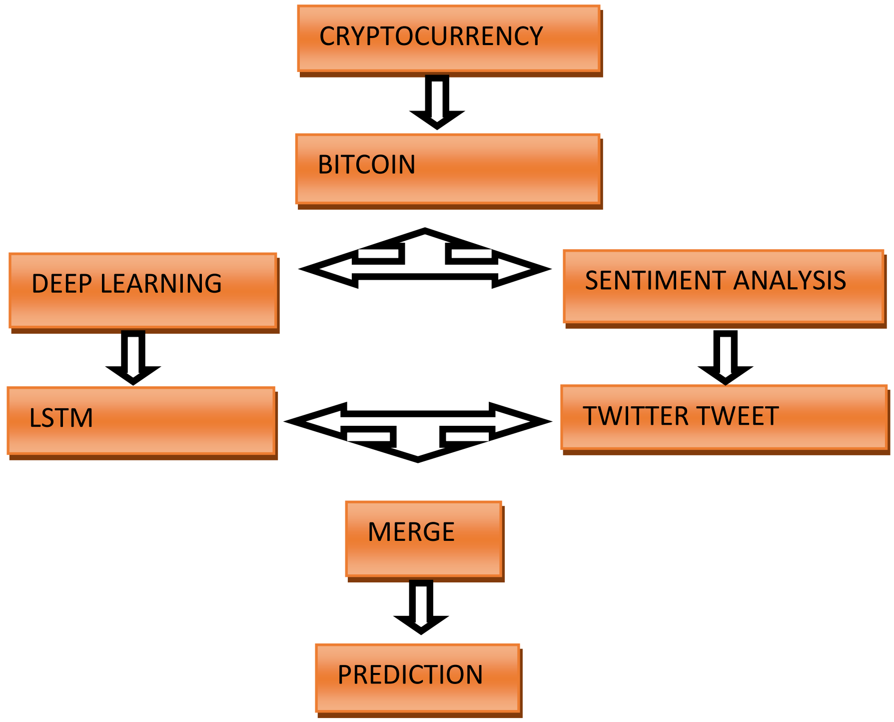
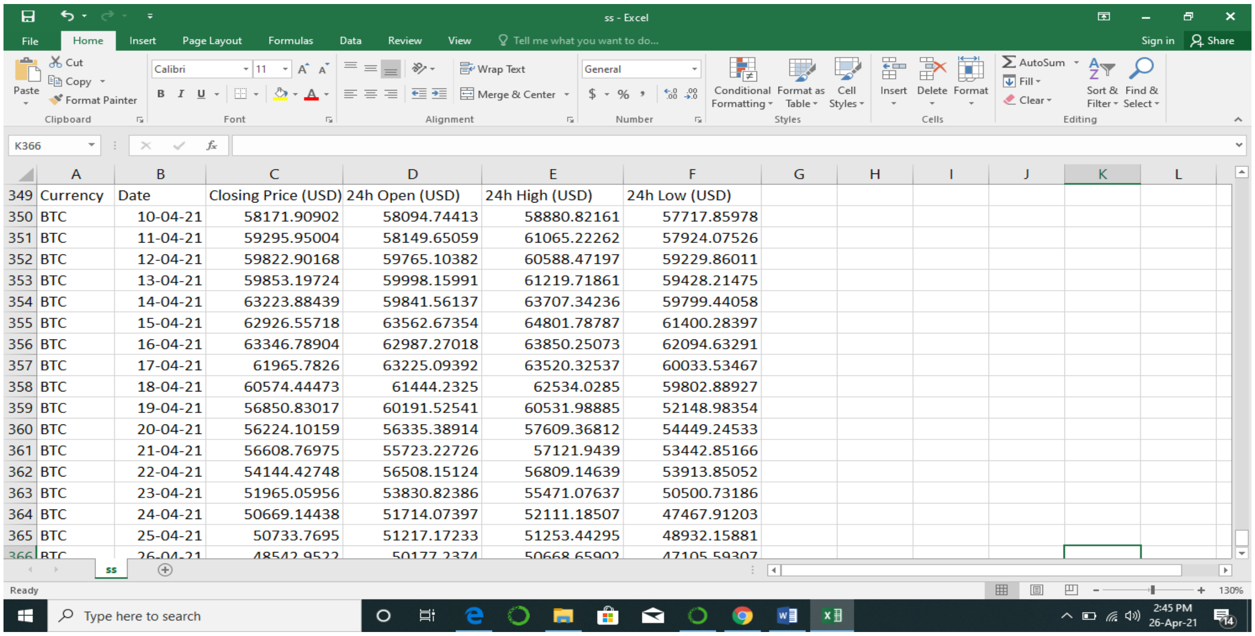


Figure 1 Workflow Diagram

DATASET

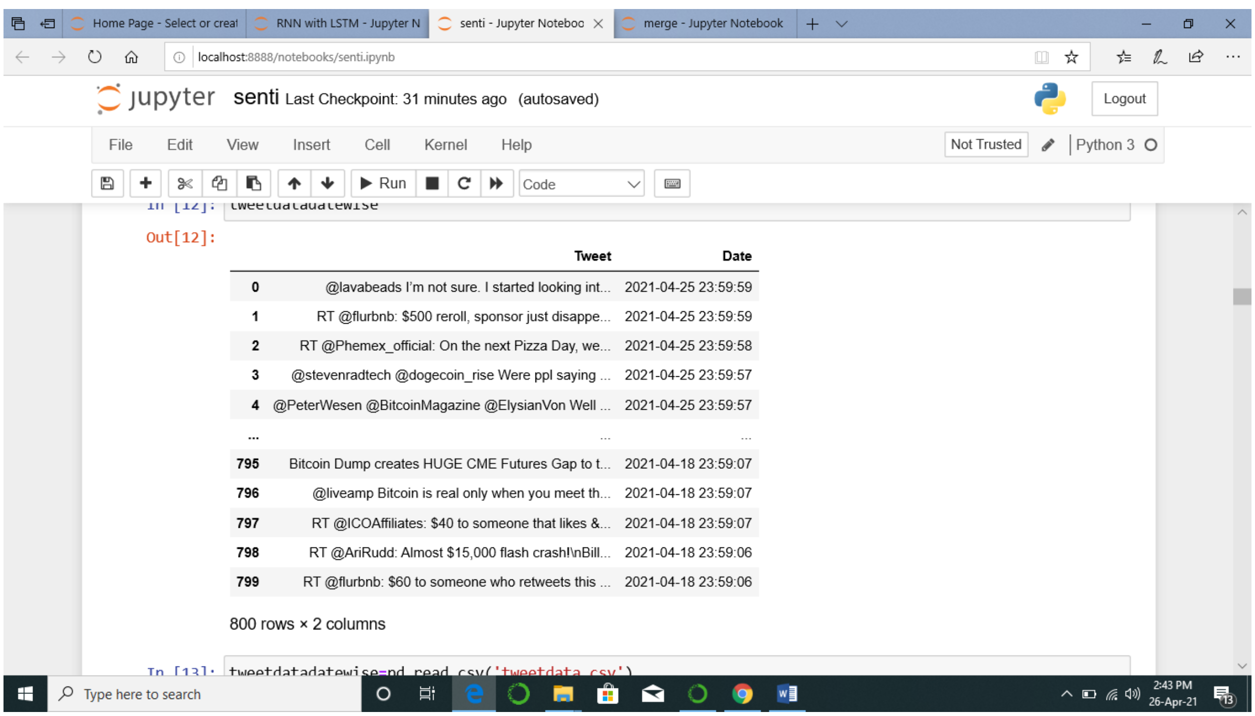
(I) BITCOIN DATA – This is the bitcoin price dataset (Figure 2) which is taken from coinbaseAPI website**.**



	Currency	Date	Closing Price (USD)	24h Open (USD)	24h High (USD)	24h Low (USD)
350	BTC	10-04-21	58171.90902	58094.74413	58880.82161	57717.85978
351	BTC	11-04-21	59295.95004	58149.65059	61065.22262	57924.07526
352	BTC	12-04-21	59822.90168	59765.10382	60588.47197	59229.86011
353	BTC	13-04-21	59853.19724	59998.15991	61219.71861	59428.21475
354	BTC	14-04-21	63223.88439	59841.56137	63707.34236	59799.44058
355	BTC	15-04-21	62926.55718	63562.67354	64801.78787	61400.28397
356	BTC	16-04-21	63346.78904	62987.27018	63850.25073	62094.63291
357	BTC	17-04-21	61965.7826	63225.09392	63520.32537	60033.53467
358	BTC	18-04-21	60574.44473	61444.2325	62534.0285	59802.88927
359	BTC	19-04-21	56850.83017	60191.52541	60531.98885	52148.98354
360	BTC	20-04-21	56224.10159	56335.38914	57609.36812	54449.24533
361	BTC	21-04-21	56608.76975	55723.22726	57121.9439	53442.85166
362	BTC	22-04-21	54144.42748	56508.15124	56809.14639	53913.85052
363	BTC	23-04-21	51965.05956	53830.82386	55471.07637	50500.73186
364	BTC	24-04-21	50669.14438	51714.07397	52111.18507	47467.91203
365	BTC	25-04-21	50733.7695	51217.17233	51253.44295	48932.15881
366	BTC	26-04-21	48542.9522	50177.2374	50668.65902	47105.59307

Figure 2 Bitcoin dataset

(II)TWITTER TWEET –This is the tweet dataset (Figure 3) which is taken from coinsbaseAPI.



	Tweet	Date
0	@lavabeads I'm not sure. I started looking int...	2021-04-25 23:59:59
1	RT @flurbnb: $500 reroll, sponsor just disappe...	2021-04-25 23:59:59
2	RT @Phemex_official: On the next Pizza Day, we...	2021-04-25 23:59:58
3	@stevenradtech @dogecoin_rise Were ppl saying ...	2021-04-25 23:59:57
4	@PeterWesen @BitcoinMagazine @ElysianVon Well ...	2021-04-25 23:59:57
...	...	...
795	Bitcoin Dump creates HUGE CME Futures Gap to t...	2021-04-18 23:59:07
796	@liveamp Bitcoin is real only when you meet th...	2021-04-18 23:59:07
797	RT @ICOAffiliates: $40 to someone that likes &...	2021-04-18 23:59:07
798	RT @AriRudd: Almost $15,000 flash crash!\nBill...	2021-04-18 23:59:06
799	RT @flurbnb: $60 to someone who retweets this ...	2021-04-18 23:59:06

800 rows × 2 columns

Figure 3 Twitter dataset

SCREENSHOTS

C. SCREENSHOTS

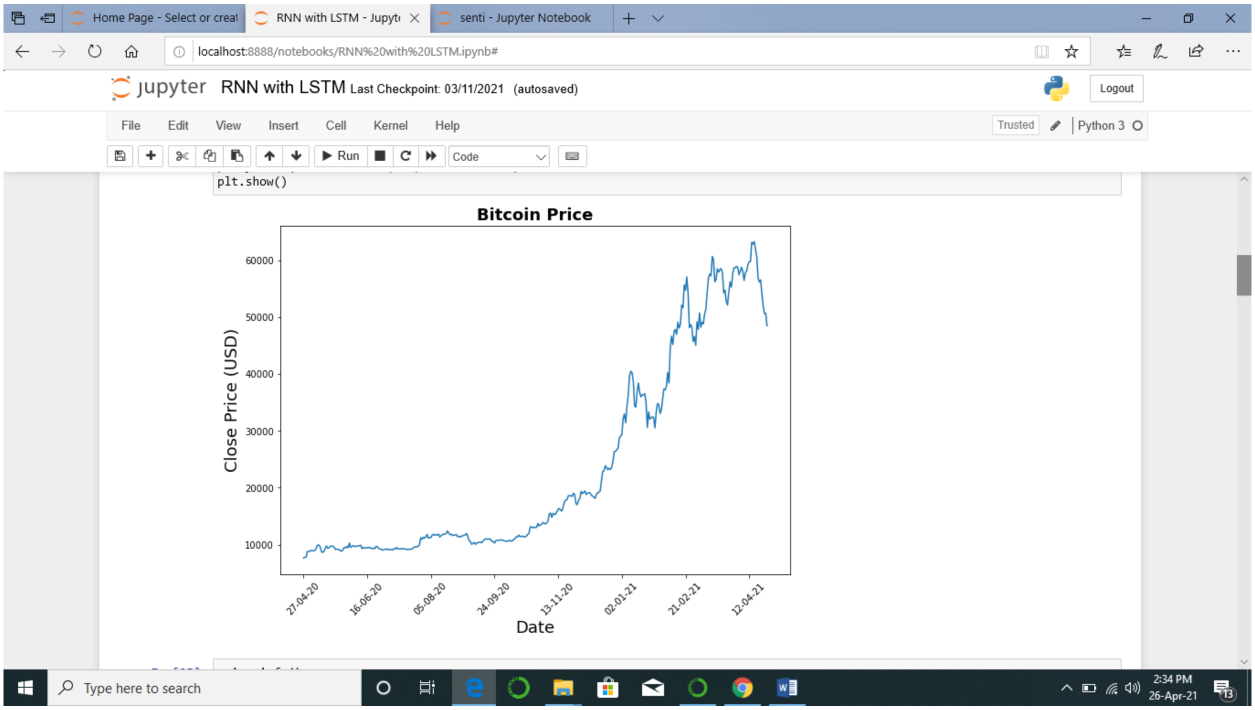


Figure 4 Dataset Visualization Plot

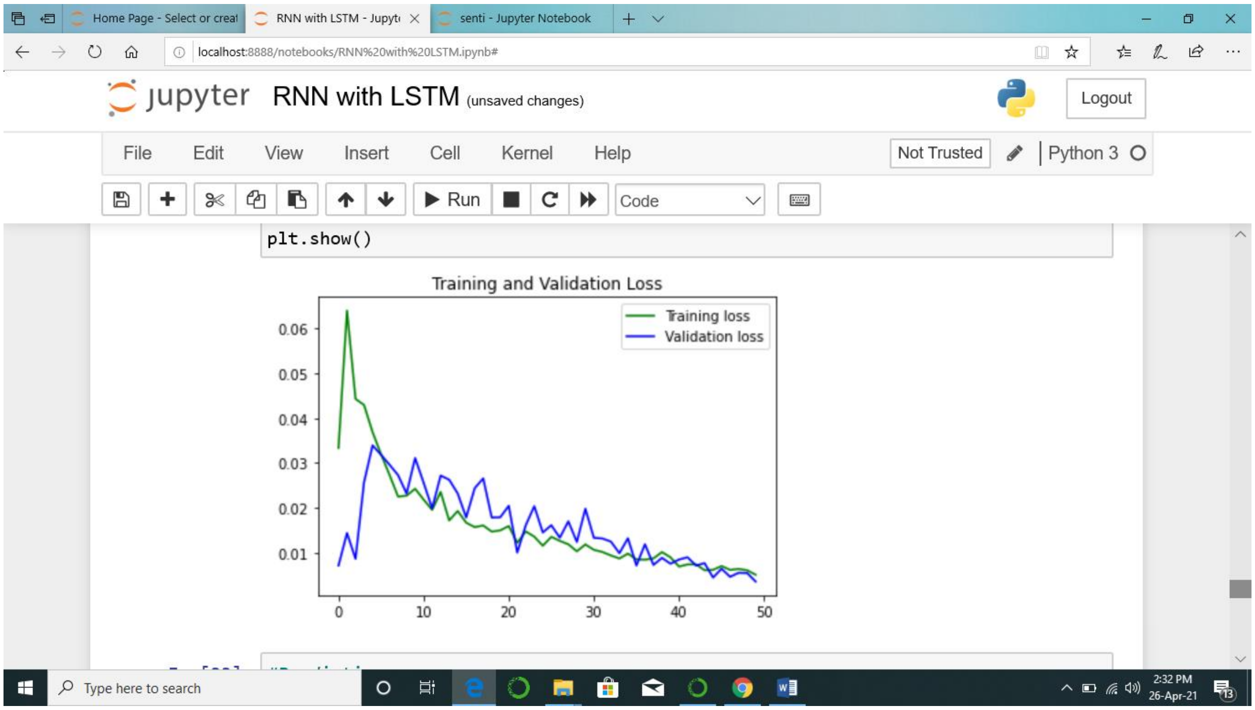


Figure 5 Training and validation plot

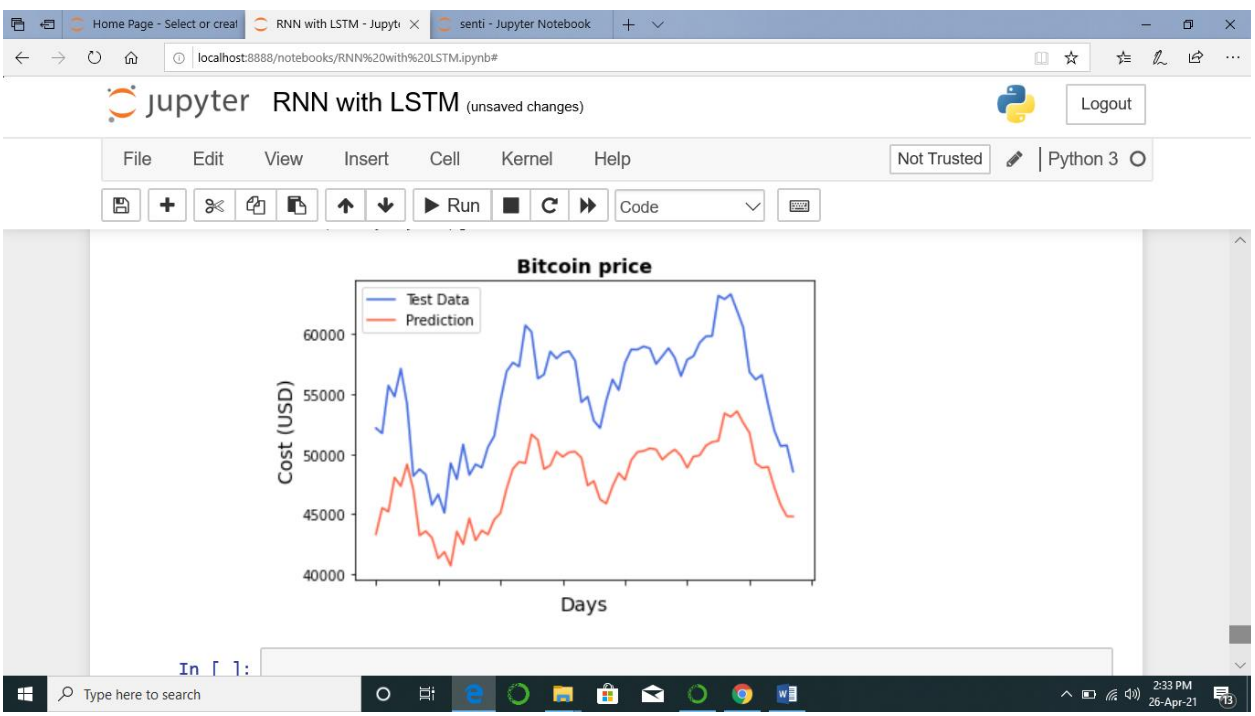


Figure 6 Bitcoin Prediction Plot

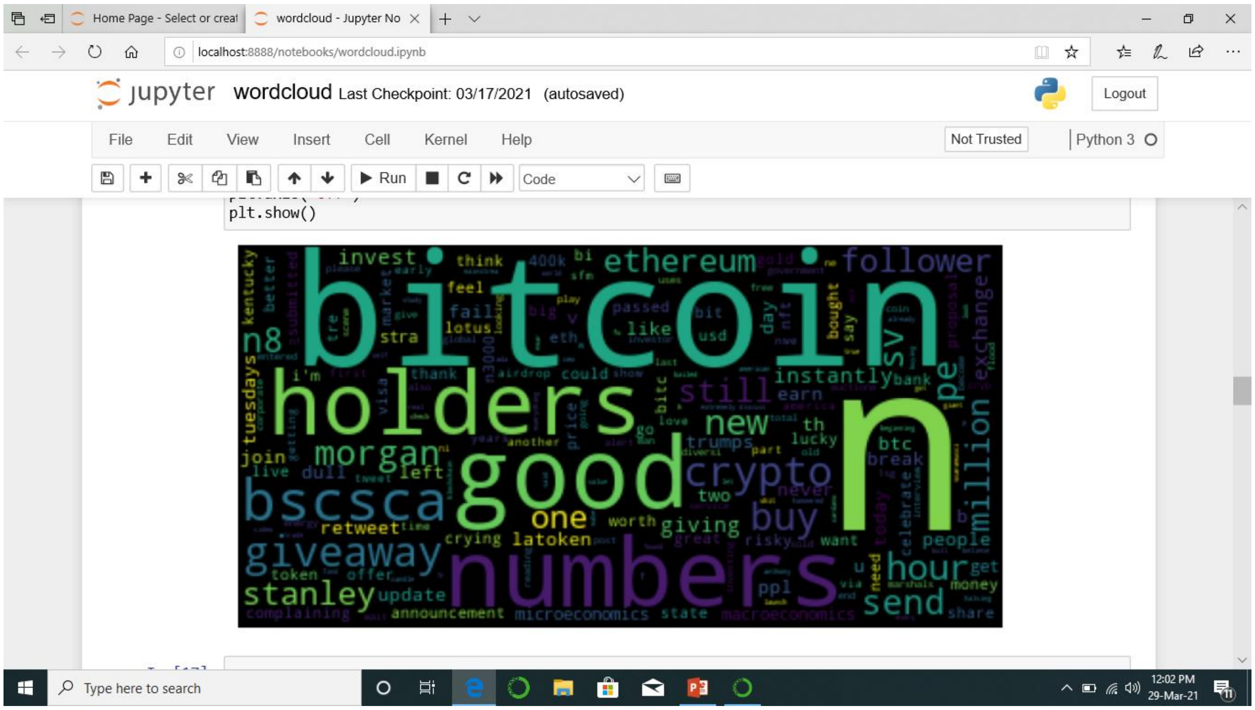


Figure 7 Twitter Tweet (Word cloud)

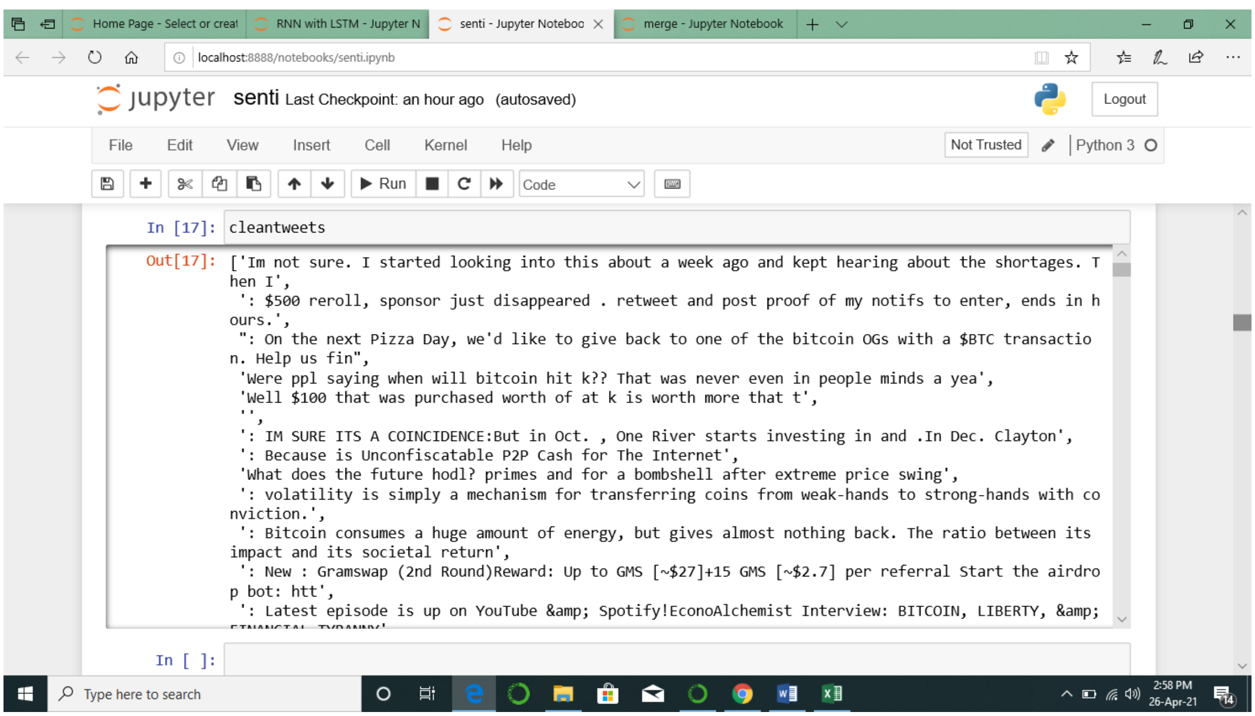


Figure 8 Twitter Tweet

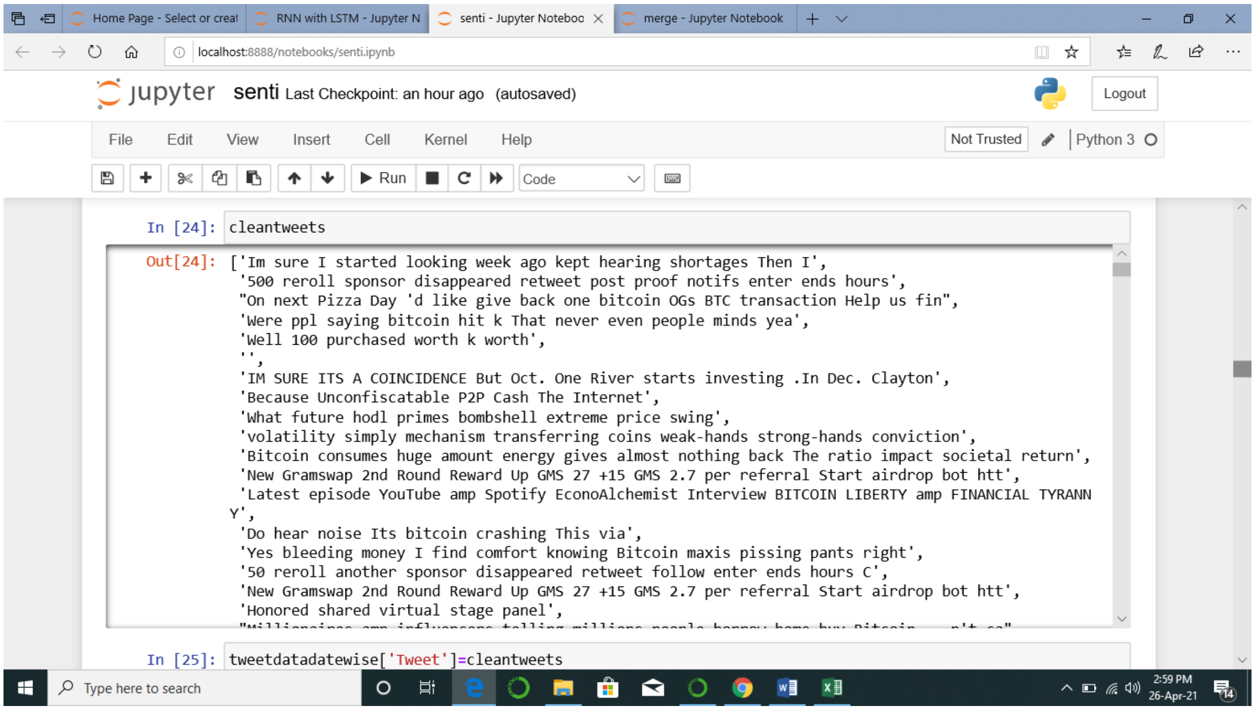


Figure 9 Twitter cleaned tweet

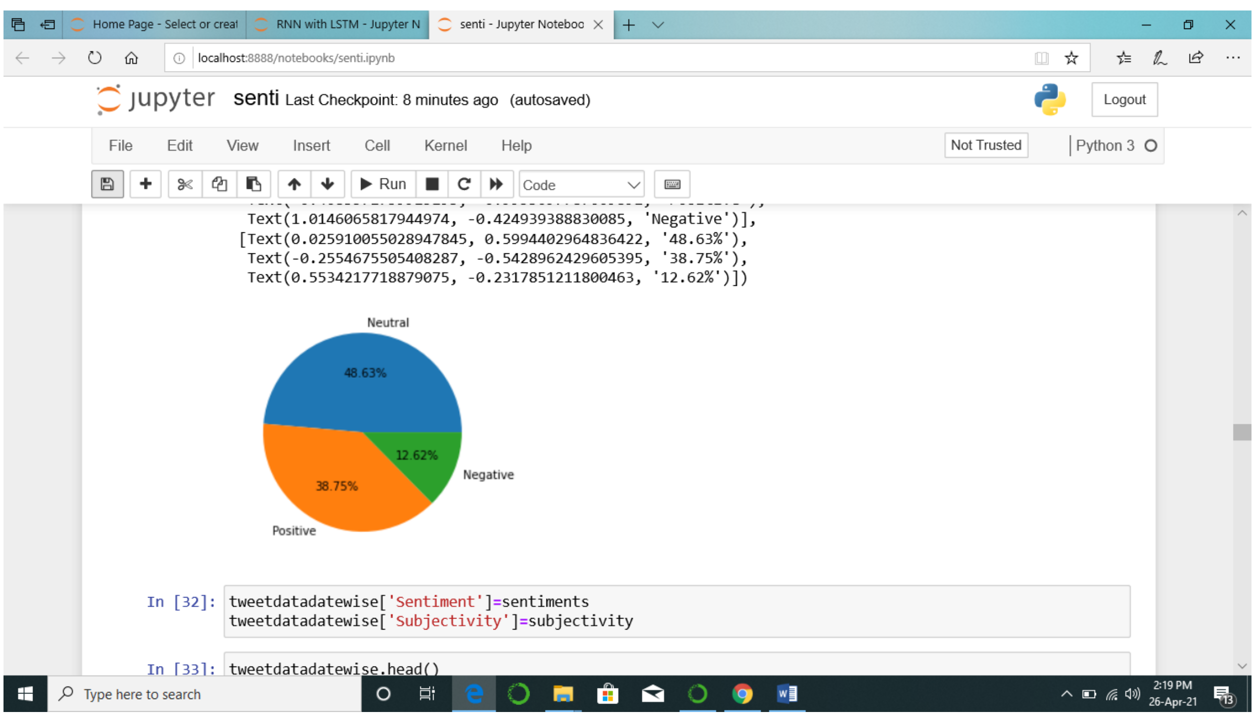


Figure 10 Tweet Polarity

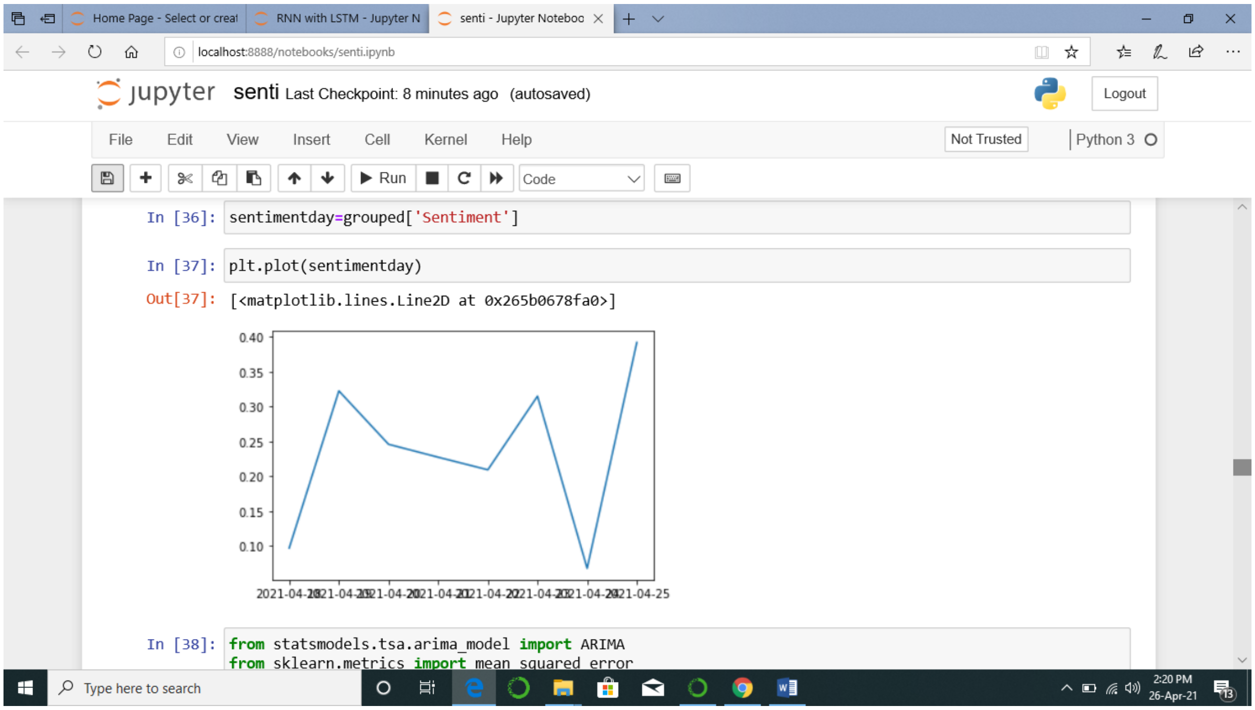


Figure 11 ARIMA Model visualization

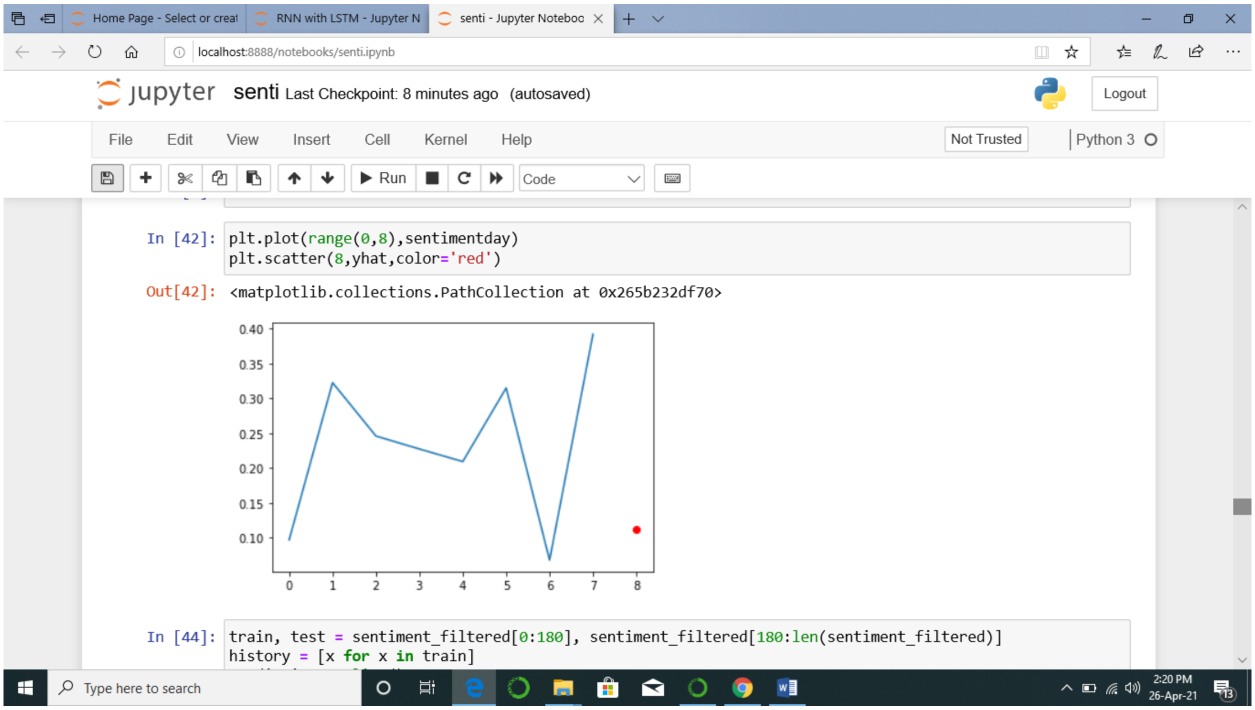


Figure 12 ARIMA Model Prediction

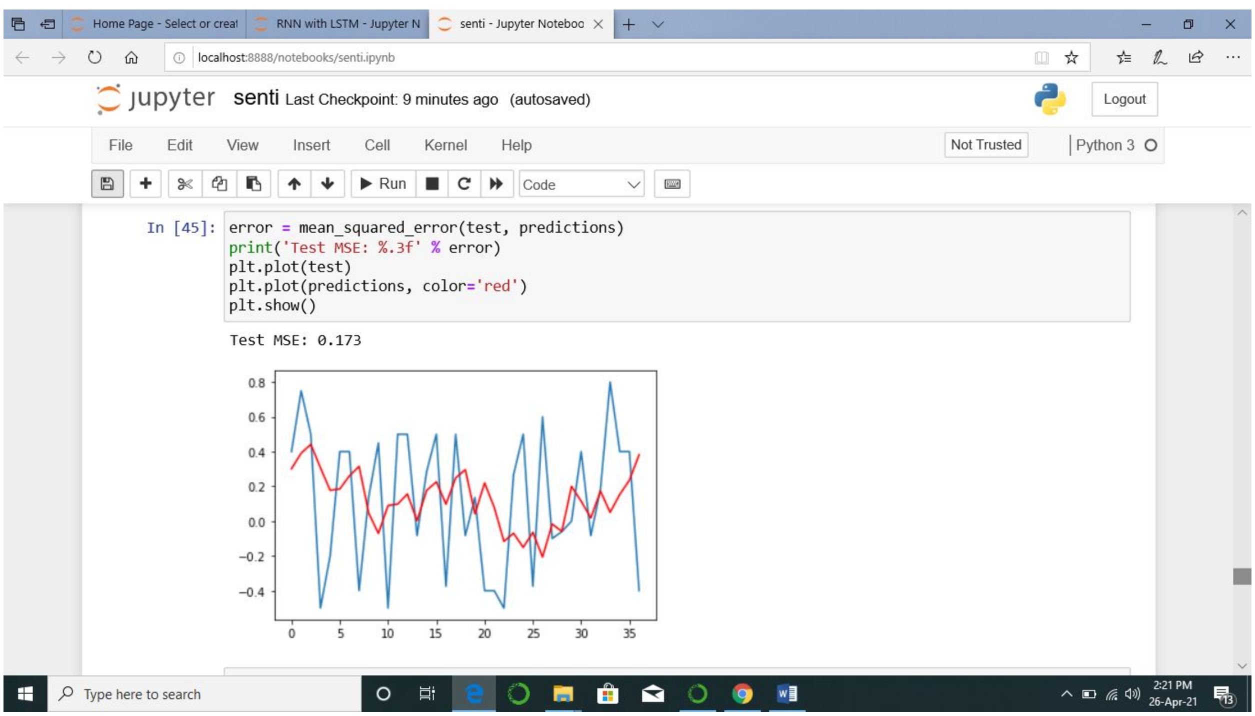
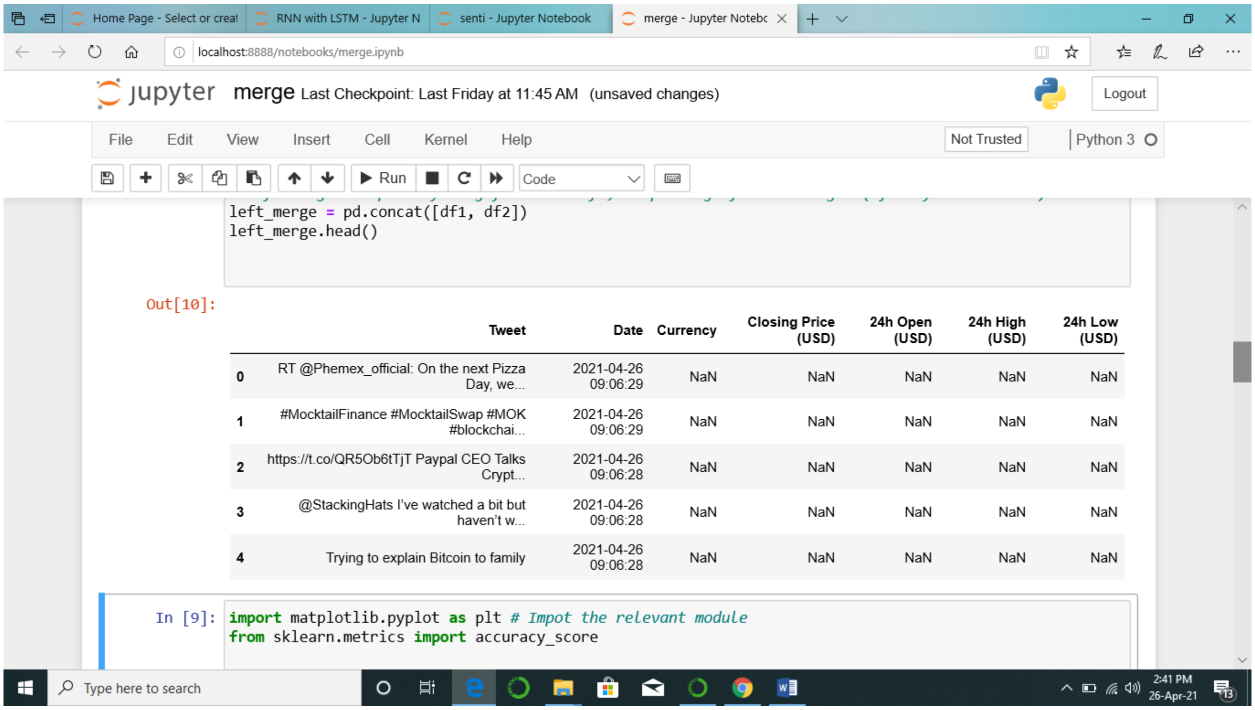


Figure 13 Mean Squared Error



```
left_merge = pd.concat([df1, df2])
left_merge.head()
```

Out[10]:

	Tweet	Date	Currency	Closing Price (USD)	24h Open (USD)	24h High (USD)	24h Low (USD)
0	RT @Phemex_official: On the next Pizza Day, we...	2021-04-26 09:06:29	NaN	NaN	NaN	NaN	NaN
1	#MocktailFinance #MocktailSwap #MOK #blockchai...	2021-04-26 09:06:29	NaN	NaN	NaN	NaN	NaN
2	https://t.co/QR5Ob6tTjT Paypal CEO Talks Crypt...	2021-04-26 09:06:28	NaN	NaN	NaN	NaN	NaN
3	@StackingHats I've watched a bit but haven't w...	2021-04-26 09:06:28	NaN	NaN	NaN	NaN	NaN
4	Trying to explain Bitcoin to family	2021-04-26 09:06:28	NaN	NaN	NaN	NaN	NaN

In [9]:

```
import matplotlib.pyplot as plt # Impot the relevant module
from sklearn.metrics import accuracy_score
```

Figure 14 Meeged Dataset

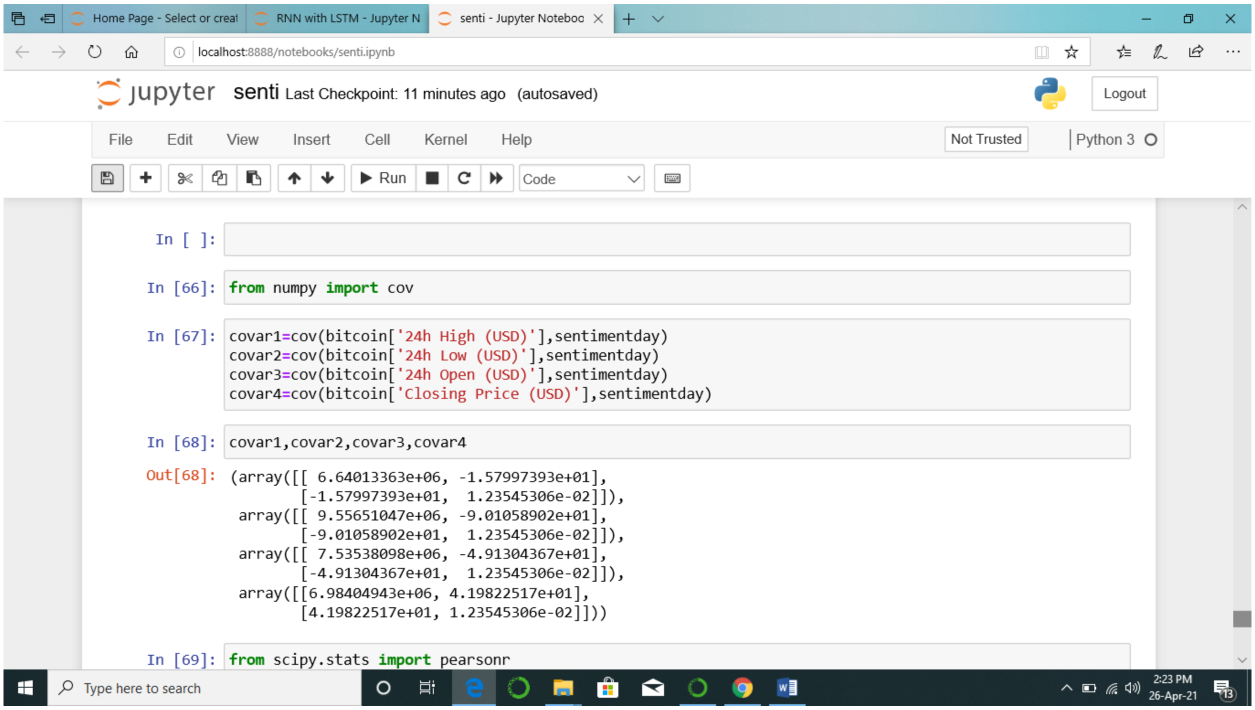


Figure 15 Co-variance correlation

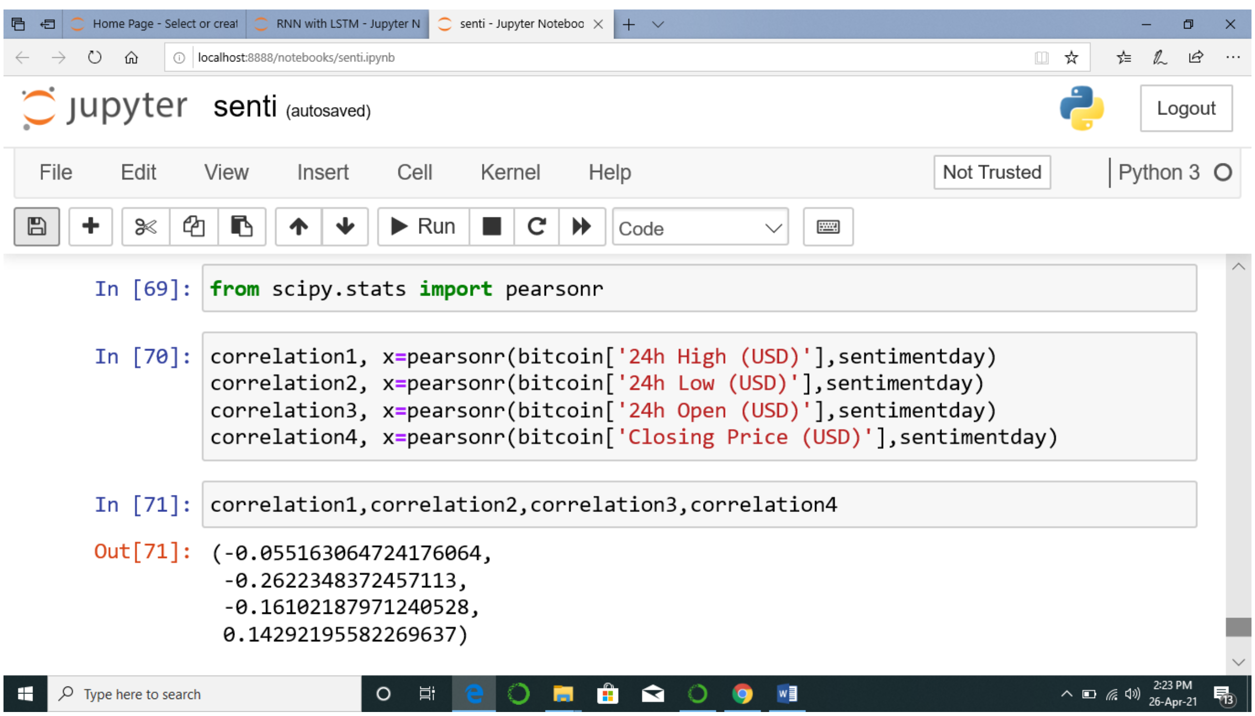


Figure 16 Pearson Correlation

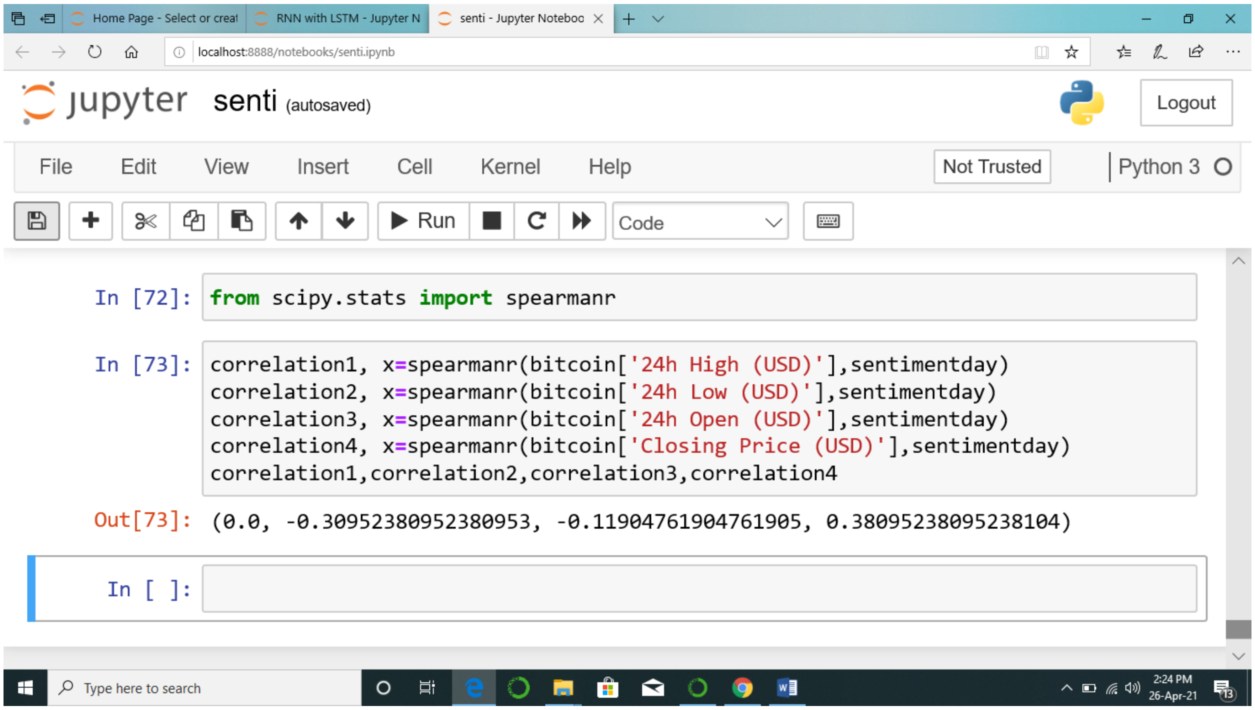


Figure 17 Spearman Correlation