It contains scripts to:

- Search for tweets related to Bitcoin.
- Preprocessing of the target tweets.
- Create different machine learning classification models.
- Aggregate the dataset into hourly intervals.
- Exploratory analysis showing various plots to show, for example, the difference in price of Bitcoin against the sentiment of tweets.

Before running the code we must ensure the following is satisfied:

- This repo assumes you have the tools I used installed
- Navigate to Google Drive
- Upload the “Dataset” folder to Google Drive
- Navigate to Google Colab
- Upload all the files from the zip containing .ipynb into Google Colab
- If you want to use the .csv file provided containing 118,596 tweets then ignore executing the file “1_Collect_Tweets.py”.

Note: please change the directory path to where you saved the dataset files in the Google Drive. The screenshot below is an example of connecting to the Google Drive directory path in Google Colab.

```
from google.colab import drive
drive.mount('/content/drive/')
%cd /content/drive/My Drive/Dissertation/
```

How to run the code:

1. Assuming you desire to use the dataset of collected tweets (ignoring having to execute the file “1_Collect_Tweets.py”), then run the file “2_Preprocess_tweets.ipynb” in Google Collab.
2. Next, run the file “3_Train_Naive_Bayes.ipynb” in Google Collab.
3. Run the file “4_Train_Logistic_Regression.ipynb” in Google Collab.
4. Run the file “5_Train_LSTM.ipynb” in Google Collab.
5. Run the file “6_count_sentiments_per_hour.ipynb” in Google Collab.
6. Run the file “7_Exploratory_analysis.ipynb” in Google Collab.

Machine Learning Methods Used:

- Naive Bayesian Network
- Logistic Regression
- LSTM Network

Tools Used:

- Botometer
- Sklearn
- Tensorflow
- Tweepy
- NLTK
- Matplotlib
- Pandas
- Seaborn

Access My repository via Github:
https://github.com/Shankar-Pankhania/Twitter_Sentiment_Analysis_on_Bitcoin