# Problem Definition

Bitcoin is one of the most trending crypto currency due to its price fluctuation phenomena. On the other hand, scientist now recognizing the power of predicting the sentiments base on the tweets for different events, political crisis and specially for economy. The sentiment of tweets on Tweeter about crypto currency directly or indirectly recognize the overall behavior of Bitcoin. In the proposed work, we will train a deep learning model for the sentiment prediction of tweets about Bitcoin. The model will take the tweets and predict their sentiment to analyze future behavior of Bitcoin.

# Dataset

## Introduction:

In the proposed work, we will use the BTC tweets sentiments dataset from Data-world platform. The selected dataset is based on the tweets of different users along with their sentiments. BTC tweets sentiments dataset is generated by collecting the tweets about Bitcoin. Here we will use the BTC dataset for the prediction of tweets sentiment by using the deep learning model.

The original Dataset contains the following number of rows and Columns:

|  |  |  |
| --- | --- | --- |
| **BTC Tweets Sentiments** | **Number of Rows in Dataset:** | 50852 |
| **Number of Variables in Dataset:** | 10 |

# Dataset – Summary of Attributes

The BTC dataset is a tabular dataset that is store in csv file format. The chosen dataset contain the total ten features/columns. The first eight features of the BTC dataset were collected manually while the next two features (New-Sentiment-Score, New-Sentiment-State) were generated with the help of NLP model. The table described the features name, description of feature, and type of data in that feature.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Variable Name | Description | Type |
| 1 | ID | Serial Number / identification Number | Numeric |
| 2 | Date | Date of the tweet when tweet posted on tweeter. | Date |
| 3 | Tweet | Content of the tweet | Text |
| 4 | Screen-name | The name of the user who posted the tweet. | Text |
| 5 | Source | List of prominent words in tweet. | Text |
| 6 | Link | Http link of extracted tweet. | URL |
| 7 | Sentiment | Sentiment of the tweet in string categorical format. i.e., positive, negative | Categorical |
| 8 | Sent-score | Label encoded form of Sentiment feature. | Numeric |
| 9 | New-Sentiment-Score | Sentiment of the tweet in string categorical format. i.e., positive, negative by NLP model | Categorical |
| 10 | New-Sentiment-state | Label encoded form of New-Sentiment-Score feature. | Numeric |

## Target Variable Description

The BTC dataset contain the sentiment and Sent-Score that have the most suitable target variables for our proposed solution. Both features contain the sentiment of tweet in string or numeric format respectively. After the initial understanding of the dataset and by considering our problem statement, we finalize the sentiment feature as our target variable. The target variable contains the three unique values that consider as classes for the prediction of sentiment.

|  |  |
| --- | --- |
| Features | ID, Date, Tweet, Screen-name, Source, Link, |
| Target | Sentiment |
| Classes | Positive, Negative, Neutral |

# Research Questions

The main objective of the proposed research is to predict the sentiment of tweet about bitcoin. There are many published studies that predict the sentiment for bitcoin analysis. But the main challenge is to gain the integrity and authenticity in prediction. On the basis of prediction result, someone will buy or sale the bitcoins. How we can achieve the better accuracy and authenticity by our prediction model is main objective of the proposed work.