

## WowSo Twitter Classification Dataset

Please note that this dataset is privately used by MagniMind bootcamp participants, solely for educational purposes. It is not publicly available.

### Data Location:

Data is stored in the mongodb. You are provided with a zip file of some data. `tweet_data.csv.zip`

### Details:

The twitter data was edited/processed by human editors, and extra fields are added. It is not in raw format.

In this project, the fields of interest to you would be:

**text:** The text field of the twits.

**topicName:** Predefined category of the twit as a string. Some of the topicNames are "Architecture", Animal, Business, Cryptocurrency, Design & Architecture, Editorial, Fitness, Interesting, Lifestyle, Memes, Nature, News, Parenting, Personal Development, Photography, Science, Startup, Social media, Technology, Travel. The categories are subject to change as the application is still evolving.

In this project, an ML classifier for the topicName: "Business" should be developed.

**edInput:** An integer in [1, 2]. This is the label in this classification problem.

1 means the classification is True.

2 means the classification is False.

-1 means, this twit is not labeled yet.

**crDate:** creation date. You will use this field to filter the twits within a timeframe, but it does not need to be in your dataframe.

Please load the data in a data frame and start exploring the data.