



DATA SCIENCE: TWITTER SENTIMENT

Predict the sentiment of the Tweet – either Positive, Negative or Neutral.

Train.csv is supplied to you.

For this week, Exploratory Analysis is not required. It will be all about Model building. The model would be evaluated based upon the accuracy of your classifier.

The expected output would be a

1. model pickle file (<https://www.geeksforgeeks.org/saving-a-machine-learning-model/>)
2. A document on the best model you obtained with the hyper-parameters
3. Code for prediction (prediction.py) on new unseen dataset; which would be exactly same as your training code with the same pre-processing steps and so on.
4. The unseen dataset (validation.csv) will be provided to you later stage and will only be used to evaluate the accuracy.
5. The prediction code would accept as input a validation.csv (same structure as train.csv) file and output the accuracy along with the confusion matrix. You are not allowed to re-train and/or tune hyperparameters based upon the validation.csv prediction results. You can do all the training, hyper-parameter tuning etc. with train.csv supplied.
6. Kindly work as a team and **be a Leader**, help your team members who are new to Text Mining.

All members of the group would have to collaborate on their own without the presence of their mentor to discuss the problems faced by each member in completing the weekly task and how they overcame them. The idea here is that the group members discuss and help each other in overcoming some of the problems they are facing individually.

The **purpose** of this group exercise is to:

- a) Imitate real world scenario where you will be working in teams/groups to collaborate and successfully deliver projects.
- b) To help you enhance some of the **most important skills required for successful careers**, like: Networking, Team-work, Communication, Group problem-solving & Leadership.

**** Disclaimer:** The dataset for this competition contains text that may be considered profane or offensive.