Guideline / Setup Instructions:

1. All the Notebooks, Data - Sets and Video Recording is uploaded on the google drive link given below:

<https://drive.google.com/open?id=1g0yMWY62tMLy6NMujOmw7kLSliKdxvlO>

2. Under DataSets:

* + *Projects + Schools + Donors + Donations:* Actual DataSets provided by DonorsChoose.org
  + *Sampled\_10K\_Records.pkl* contains sampled 10000 records used for Clustering
  + *DownSampled\_Balanced\_Data.pkl* is the data file with balanced classes used for Classification
  + *Data Enrichment 1* contains dataset of Texas Housing Data
  + *Data Enrichment 2* contains dataset of School Ratings from Niche.com
  + Note: *Data Enrichment 3* is the combination of Data Enrichment 1 + 2

3. The shape of the data after merging Projects, Schools and Donations was 3732821 record with 32 columns. Hence, we have taken 10K sample records out of the entire data set to run all our algorithms. For classification, we were facing the problem of imbalanced classes so we have used down - sampling technique to take the sub samples of majority classes and have a count match between both the classes.

4. All the notebooks are also uploaded on GitHub at below link:

<https://github.com/4dataseekers/CMPE257_MachineLearning.git>

Thank you,

TEAM SEEKERS