1. Complete model evalution
2. Complete evalution pipeline
3. Run training.py as docker image
4. Run prediction.py as docker image
5. Run evalution.py as docker image
6. Create api for training, prediction, evalution
7. Deploy code in cloud

Load data

1. Currently we are using load\_data.py where to get data from different sources we have written methods, we are importing this python file
2. In next iteration, change load data in such a way so that you get data from any source by calling one method and the source name can be provided by config file, or properties file

Question:

1. Why not to give absolute path rather than ..
2. Why do we do feature extraction and EDA in the era of end to end machine learning
3. can we run a python Jupyter notebook with two kernel, half code with one kernel and another half with another kernel
4. why we import numpy as np in different files, cannot we use one centralize np

imp

what is the difference between gradient boost and xgboost, random forest, adaboost

feature selection

select columns which does not have correlation

todo

1. training JN cleaning
2. scoring JN cleaning
3. logging functionality
4. artifacts in git repo