System: OS X, Linux, Windows 10
Language: Python 2.7.10
External Libraries: Numpy, Pandas, Sklearn
Tool: Pycharm 5.0.1

Files (folders):
1) readme - contains instructions to compile and run programs.

2) source_code - contains Python program for all tasks and needed files.

dataloader.py: Iuput .csv data from current file fold

writetxt.py: Write a data into a .txt file which save some data we need to use in future

datawriter.py: Output data to current folder

dataprocessing.py: This file would do process the data from BNP scoure, The data-preprocessing method include that deal with missing value, factorize categorical value, convert type of data, shuffle and sort the data, as well as increasing and reducing the dimension of the dataset

statistic.py: This part is to analyze statistically the relevance of each feature and target. the distribution of target and each feature,especially categorical feature. Based on this statistic output, we can have the weight of each value in each categorical feature, the distance between each datapoint. Deeply, We can even guess some meaning of some features.

Paint_statistic.py: This part is to visualize the dataset, by using functions in this module, We will have the graph of relation of each feature, the value distribution of each feature, relation of target and each feeture, the distribution of target as well as weight of each feature.

Classification.py: In this part, We have 6 classification methods to train the model including Xgboost, Extremely random tree, SVM, Knn, naive bayes and logistic regression. we will have the logloss or accuracy rate as to evaluate the how good about the model is.
Also, We plot the ROC and AUC in each method in this module.

Main.py: This is the main module

3) report:
1. Title with group information
2. Abstract
3. Introduction of the background
4. Problem definition and formalization
5. Data description and preprocessing
6. Methods description (detailed steps)
7. Experiments design and Evaluation
8. Conclusion
9. References

4) task distribution form – contains the task of each person in our group

How to install libraries:

1) For numpy, type " sudo apt-get install python-numpy" in terminal.

2) For sklearn, If you already have a working installation of numpy and scipy, the easiest way to install scikit-learn is using pip install -U scikit-learn

3) For Pandas, you can download from this website “http://pandas.pydata.org/ ”.