Data Science: A Programming Approach
Mahyar S Vaghefi
University of Texas Arlington

Individual Project #2

This project involves two separate mini-projects.

Titanic-Machine Learning from Disaster

- **Step 1.** Join kaggel at https://www.kaggle.com/ , if you have not already done that.
- **Step 2.** You need to compete in the ***Titanic Project*** (accessible at https://www.kaggle.com/c/titanic)
- **Step 3.** Use train.csv file (available at kaggle web site) and predict unseen data on test.csv (available at kaggle web site).
  - Read the full desciption of the competition on the website
  - Download train and test datasets and load them into your environment (Jupyter)
  - **Survived** is your target (dependent variable)
  - You can use any combinations of available features to develop your models (Keep in mind that there are missing values is some features)
  - Use different classification methods (the ones that you learned in this course) to create your predictive models.
    - For each method create a separate Jupyter file
  - After developing your model(s), use them to predict the test data. Then, save your result as a separate CSV file.
    - You can find a sample of submission file in **"Data"** tab inside the project web site
  - One member of the group needs to submit the final prediction result (in a defined format) to kaggle. you can also creat a team and submit it as a team.
  - Capture your score. You need to report it.
- **Step 4.** Add the score for each of the prediction models at the end of the Jupyter file (as a comment).
- **Note.** DO NOT use other people code. The main idea is to use your own knowledge that you gained in the course. If I notice that any individual use outsider codes, then you will get no point for this part of the project.

Output

- Make sure to put descriptive comments on your code
- Report your result in a separate Word document
- Put your code files, and analysis report in a zip folder and submit it in Canvas

Due Date: May 7 2020 at 7:00 PM

Grading Criteria

Comprehensiveness	**30%**
Correctness	**20%**
Complete Report	**20%**
Clear Code	**20%**
Innovation (Extra)	**20%**
Total	**120%**