Homework-6

**Out Date:** 10/18/2019 (Friday)

**Due Date:** 10/31/2018 (Thursday) 11:59PM

Team#: \_6\_\_

Team Member-1: Thanh Hung Duong\_\_\_\_\_\_\_\_\_\_\_\_Member’s Contribution (in %) 50

Team Member-2: Johnny Nino Ladino\_\_\_\_\_\_\_\_\_\_\_Member’s Contribution (in %) 50

**Submission**

1. Work on the Problme-1.
2. Prepare your Python file for Problem-1 (e.g., HW6\_P1\_Team#.py).
3. Upload the files to blackboard.

**Problem Statement:** This is a binary classification problem, where the goal is to classify each passenger of the Titanic ship into survived (1) or not survived (0) groups. The training and test datasets are available. Your task is to develop OneR models (from the training set) to predict which passengers (in the test set) survived the tragedy.

OneR, short for "One Rule", is a simple, yet accurate, classification algorithm that generates one rule for each feature in the data. There are 7 features in the Titanic dataset (item-1 to item-7 in the following list) and one target class (item-8) as shown below:

1. **pclass:** Ticket class (1 = 1st class, 2 = 2nd class, 3 = 3rd class)
2. **gender:** Gender of the passenger (Male, Female)
3. **age:** Age in years
4. **sibsp:** # of siblings or spouses aboard the Titanic
5. **parch:** # of parents or children aboard the Titanic
   1. Parent = mother, father
   2. Child = daughter, son, stepdaughter, stepson
   3. Some children travelled only with a nanny, therefore parch=0 for them.
6. **fare:** Passenger fare
7. **embarked:** Port of Embarkation (C = Cherbourg, Q = Queenstown, S = Southampton)
8. **survived:** Survival (0 = Not survived, 1 = survived). This is a binary class.

OneR essentially is a majority voting. For instance, if the majority of the female passengers (gender = female) in the training set survived the tragedy, the OneR classifies all the females in the test set as survived (1). If the majority of the male passengers (gender = male) in the training set did not survive the tragedy, the OneR classifies all the males in the test set as not survived (0). Therefore, OneR for this feature could be the following:

OneR(gender) = *if the passenger was female, she survived.*

Similarly, if the majority of the first class passengers (pclass = 1) in the training set survived the tragedy, the OneR classifies all the first class passengers in the test set as survived. Next, its checks the second class passenger (pclass=2). If the majority of the second class passengers (pclass = 2) in the training set survived the tragedy, the OneR classifies all the second class passengers in the test set as survived. Finally, repeat this step for the third class passengers (pclass = 3). Let’s assume that the majority of the third class passenger in the training set did not survived. In this case, OneR for this feature could be the following:

OneR(pclass) = if *the passenger travelled in the first class or the second class, s/he survived*.

Develop a Python script to

1. Read-in the training set and test set.
2. Construct a machine learning model using OneR for each of the following features: gender, pclass, sibsp, parch, embarked. **[50 points; 10 points per feature]**
3. For each feature, predict *class label* (survived or not survived) for every passenger in the test set. **[10 points]**
4. For each feature, store the passenger IDs and class labels in the appropriate worksheet (e.g., Gender\_Based\_Prediction for gender based OneR) of the attached output file (titanic\_test\_predictions.xlsx). **[10 points]**
5. Compute success rate for the test set. Success rate is the total number of correct predictions divide by the total instances. **[10 points]**

Use pandas, and numpy for data analysis. Submit the Python script to blackboard. The script should be self-contained, meaning it should read the data files (titanic\_traning.xlsx, and titanic\_test.xlsx) from the folder where the script is and produce the final results (titanic\_test\_predictions.xlsx).

Please make sure your code follows the Python programing style guide available here: <https://www.python.org/dev/peps/pep-0008/> **[10 points]**.

Please make sure the code is well-commented **[10 points]**