الاسم: يوسف علاء الدين شاكر

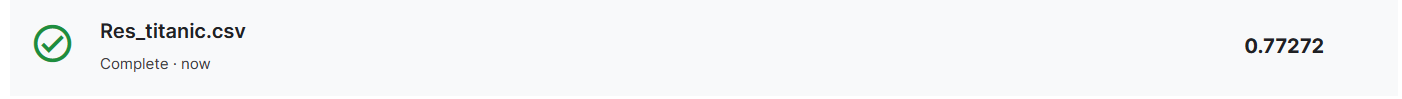
رقم الجلوس: 20191700920

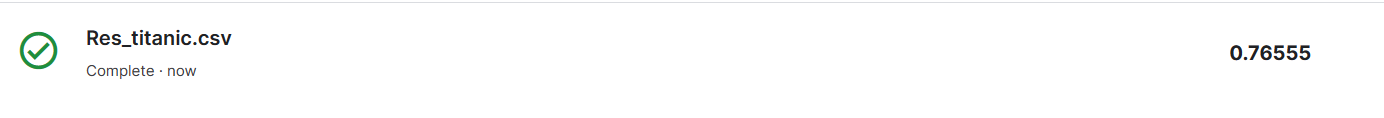
الفرقة: الرابعة

CS - sec 6القسم:

# First Trail

* Preprocessing
  + Target column is “Survived”
  + First time using :
    - Columns Features: "Sex","Age","SibSp","Parch"
    - Only Age has NA
    - Replace NA values with column median
    - Change category data in Sex with (0) for female and (1) for male
    - No values from any columns were normalized
* Accuracy
  + Got accuracy 77% from Kaggle submissions using Decision Trees and Naïve bayes



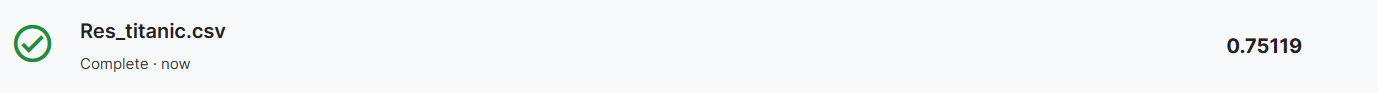
* + Got accuracy 76% from Kaggle submissions using Logistic Regression 
  + Got accuracy 76% from Kaggle submissions using SVM with linear kernel and type is 'C-classification' .

# Second Trail

* Preprocessing
  + Target columns is: “Survived ”
  + Second time:
    - Drop: Passengersid, Name,Ticket ,Cabin
    - Normalize fare column
    - Change categorical data:
      * Sex female 0 and male 1
      * Take-off for S = 1 , C = 2, Q = 3
    - Replace NA values in age with mean value of the column
* Accuracy
  + Got accuracy 76% from Kaggle submissions using Logistic Regression



* + Got accuracy 75% from Kaggle submissions using naïve bayes and SVM with radial kernel and type is 'C-classification' .



* + Got accuracy 78% from Kaggle submissions using Decision trees .

