Statistical and Predictive Modeling II (DATA 2204)
Assignment #2 – Logistical Regression (15% of Final Grade)
Professor: Sam Plati

Mr. John Hughes would like you to create a ***Logistical Regression*** model for his **heartfailure.csv** dataset in order to predict if the patient has deceased during the follow-up period. Below are the dataset variables:

Independent Variables

age: age of the patient (years)
anaemia: decrease of red blood cells or hemoglobin (boolean)
high blood pressure: if the patient has hypertension (boolean)
creatinine phosphokinase (CPK): level of the CPK enzyme in the blood (mcg/L)
diabetes: if the patient has diabetes (boolean)
ejection fraction: percentage of blood leaving the heart at each contraction (percentage)
platelets: platelets in the blood (kiloplatelets/mL)
sex: woman or man (binary)
serum creatinine: level of serum creatinine in the blood (mg/dL)
serum sodium: level of serum sodium in the blood (mEq/L)
smoking: if the patient smokes or not (boolean)
time: follow-up period (days)

Dependent Variable

death event: if the patient deceased during the follow-up period (0-Alive, 1-Deceased)

The Ask:

1. Create a PowerPoint (PPT) presentation that includes the following:
   a. Cover Page (Title, Name (1st and last) and Student Number)
   b. Analysis Statement (i.e. create Logistical Regression Model) – *2%*
   c. Present the Learning Curve for the Standard Logistical Regression Model and explain **two (2)** insights – *2%*
   d. Present Standard and Optimized Regression outputs and explain: *- 9%*
      - Entire Classification Report (both Standard and Optimized Models)
      - ROC/AUC Curve (only for the Optimized Model)
   e. State and explain **two (2) recommendations** for Mr. John Hughes for next steps. – *2%*

Attention: Please ensure that all key facts are in your slides and not in the notes section

Hint: Leverage the code from Wk4-LogReg

Random State = 100 for all section

2. Provide an HTML copy of your python code

Please post your PowerPoint Document (.ppt) and HTML Python Code via assignments under Assignment #2 by 11:59 p.m. on Friday, June 4th, 2021

Grading Rubric

	Needs Improvement	Average	Above Average	Comments
1. Create a PowerPoint (PPT) presentation that includes the following: a) Cover Page (Title, Name (1st and last) and Student Number) b) Analysis Statement (i.e. create Logistical Regression Model) – *2%* c) Present the Learning Curve for the Standard Logistical Regression Model and explain **two (2)** insights – *2%* d) Present Standard and Optimized Regression outputs and explain: - *9%* • Entire Classification Report (both Standard and Optimized Models) • ROC/AUC Curve (only for the Optimized Model) e) State and explain **two (2) recommendations** do you have for Mr. John Hughes for next steps. – *2%* 2. Copy of your HTML Python Code				

Needs Improvement –Missing the minimum requirements stated in the assignment requirements.
Average –Meets the minimum requirements stated in the assignment requirements.
Above Average –Exceeds the requirements that are stated in the assignment requirements.