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

Mr. John Hughes would like you to create a *Logistical Regression* **model and associated ROC/AUC curve** for his **cancer.csv** dataset in order to predict if the patient has cancer based the following variables:

Independent Variables

ID - ID number
Clump Thickness - 1-10
UofCSize - Uniformity of Cell Size 1-10
UofShape - Uniformity of Cell Shape 1-10
Marginal Adhesion - 1-10
SECSize - Single Epithelial Cell Size 1-10
Bare Nuclei - 1-10
Bland Chromatin - 1-10
Normal Nucleoli - 1-10
Mitoses - 1-10

Dependent Variable

Class - Benign (i.e. No Cancer) - 2, Malignant (i.e. Cancer) - 4

Note: ID will not be used and will need to be dropped prior to building your model.

The Ask:

1. Create a PowerPoint (PPT) presentation that includes the following:
    a. Cover Page (Title, Name (1st and last) and Student Number)
    b. Rational Statement (summary of the problem or problems to be addressed by the PPT) – *2%*
    c. Present the Learning Curve for the Logistical Regression standard model and identify **two (2) insights** – *2%*
    d. Present and Explain **three (3) key insights** from the classification report metrics (i.e. Precision, Recall, F1) for the Optimized Logistical Regression Model – *7%*
    e. Present and Explain **two (2) key insights** from ROC/AUC Curve (Optimized Model) – *2%*
    f. 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 Wk4b-LogMulti
Random State = 100 for all section

2. Provide an HTML copy of your python code

Please post your <u>PowerPoint Document (.ppt) and HTML Python Code</u> via assignments under Assignment #2 by 11:59 p.m. on Thursday, February 17th, 2022

Grading Rubric				
	Exemplary (14-15)	**Proficient (10-13)**	**Incomplete (7-9)**	**Needs Improvement (0-6)**
Analysis	Cover Page Complete Rational Statement is complete with supporting details Learning Curve presented with two (2) insights presented with explanation/ justification Classification report presented with three key (3) insights presented and fully evaluated from the Optimized Logistical Regression model ROC Curve presented with two (2) detailed insights	Cover Page Complete Rational Statement is complete with high-level supporting details Learning presented with two (2) insights presented with high-level explanation/ justification Classification report presented with three key (3) insights presented with high-level evaluations from the Optimized Logistical Regression model ROC Curve presented with two (2) high-level evaluations	Cover Page Incomplete Rational Statement is complete with missing supporting details Learning Curve presented with less than two (2) insights and/or Missing explanation/ justification Classification report presented with less than three key (3) insights presented and evaluated from the Optimized Logistical Regression model ROC Curve presented with less than two (2) insights presented and evaluated	Cover Page missing Rational Statement missing Learning Curve and/or insights are missing or incorrect. Classification report and insights are missing or incorrect. ROC Curve presentation and/or evaluations are incorrect or missing
Next Steps	Two (2) recommendations have been identified with detailed explanations.	Two (2) recommendations have been identified with only high-level explanations.	Less than Two (2) recommendations and incomplete explanations.	Recommendations are missing or incorrect.

Note: 50% Grade Penalty for missing Python HTML File