Statistical and Predictive Modeling II (DATA 2204)
Assignment #3 – Discriminant Analysis (15% of Final Grade)
Professor: Ritwick Dutta

Mr. John Hughes is looking at developing an LDA model for his **cancer.csv** dataset and evaluate its effectiveness. If you recall the dataset has 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.

Below are the results of the Optimized Logistical Regression model (with SMOTE):

```
Optimized Model

Model Name: LogisticRegression(class_weight='balanced', random_state
=100)

Best Parameters: {'clf__C': 1, 'clf__penalty': 'l2'}

 [[89  0]
 [ 3 45]]

              precision    recall  f1-score   support

           2       0.97      1.00      0.98        89
           4       1.00      0.94      0.97        48

    accuracy                           0.98       137
   macro avg       0.98      0.97      0.98       137
weighted avg       0.98      0.98      0.98       137
```

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. Identify and explain **two (2) key insights** from the Pandas Profile Report – ***2%***
   d. Present and explain **three (3) key insights** from the Optimized LDA classification report**,** but first use **SMOTE** to ensure that the dataset is balanced. – ***6%***
   e. Compare the Optimized LDA to the Optimized Logistical Regression model (from page 1) identifying **three (3) key insights**. – ***3%***
   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 Wk5a-LDAQDA

Random State = 100 for all section

2. Provide a copy of your HTML Python Code

Please post your PowerPoint Document (.ppt) and HTML of Python Code via assignments under Assignment #3 by 11:59 p.m. on Thursday, March 10th, 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 Two (2) insights from Pandas Profile Report presented with explanation/ justification Classification Report and three (3) LDA Model key insights presented and fully evaluated Three (3) comparison insights presented and fully evaluated	Cover Page Complete Rational Statement is complete with high-level supporting details Two (2) insights from Pandas Profile Report presented with high-level explanation/ justification Classification Report and three (3) LDA key insights presented and with high-level evaluations Three (3) comparison insights presented with high-level evaluations	Cover Page Incomplete Rational Statement is complete with missing supporting details Pandas Profile Report presented with less than two (2) insights and/or Missing explanation/ justification Classification Report and less than three (3) LDA key insights presented and evaluated Less than three (3) comparison insights presented and evaluated	Cover Page missing Rational Statement missing Insights from Pandas Profile Report are missing or incorrect. Classification Report and LDA Model key insights missing or incorrect Comparison insights missing or incorrect
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