Statistical and Predictive Modeling II (DATA 2204)
Assignment #5 – Support Vector Machines (15% of Final Grade)
Professor: Ritwick Dutta

Mr. John Hughes would like you to revisit the **cancer.csv** dataset and create a standard and optimized SVM model. If you recall the dataset has the following variables.

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.

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 Original SVM Model and explain **three (3) insights** ***– 3%***
   d. Present and explain **three (3) key insights** from the Optimized SVM model classification report, but first **use SMOTE to balance the Classes**. ***– 7%***
   e. State and explain **three (3) recommendations** for Mr. John Hughes for next steps. ***– 3%***

Attention: Please ensure that all key facts are in your slides and not in the notes section
Hint: Leverage the code from Wk11c-SVM-Tutorial
Random State = 100 for all section

2. HTML of your Python Code

Please post your PowerPoint Document (.ppt) and HTML Python code via assignments under Assignment #5 by 11:59 p.m. on Wednesday, April 6st , 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 Three (3) insights and detailed explanations/ justifications Classification report presented with three key (3) insights presented and fully evaluated	Cover Page Complete Rational Statement is complete with high-level supporting details Learning Curve presented with Three (3) insights and high-level explanations/ justifications Classification report presented with three key (3) insights with high-level evaluations	Cover Page Incomplete Rational Statement is complete with missing supporting details Learning Curve presented with less than three (3) insights and/or Missing explanation/ justification Classification report presented with less than three key (3) insights and evaluations	Cover Page missing Rational Statement missing Learning Curve not presented and explanations/justifications are missing/Incorrect Classification Report and/or insights are missing or incorrect.
Next Steps	Three (3) recommendations have been identified with detailed explanations.	Three (3) recommendations have been identified with only high-level explanations.	Less than Three (3) recommendations and incomplete explanations.	Recommendations are missing or incorrect.

Note: 50% Grade Penalty for missing Python HTML File