Introduction to Data Analysis (DATA 1200)
Assignment #4 – SVM and Naïve Bayes (15% of Final Grade)
Professor: Ritwick Dutta

Mr. John Hughes has been thinking that maybe an **SVM and Naïve Bayes model** might produce better results for the **raisin_dataset.csv**. If you recall the dataset contains 900 observations and 8 variables:

Independent Variables

Area: Gives the number of pixels within the boundaries of the raisin.
Perimeter: It measures the environment by calculating the distance between the boundaries of the raisin and the pixels around it.
MajorAxisLength: Gives the length of the main axis, which is the longest line that can be drawn on the raisin.
MinorAxisLength: Gives the length of the small axis, which is the shortest line that can be drawn on the raisin.
Eccentricity: It gives a measure of the eccentricity of the ellipse, which has the same moments as raisins.
ConvexArea: Gives the number of pixels of the smallest convex shell of the region formed by the raisin.
Extent: Gives the ratio of the region formed by the raisin to the total pixels in the bounding box.

Dependent Variable

Class: Kecimen and Besni raisin.

The Ask:

1. **Create a Python Script using Jupyter Notebook (then convert to .html)** *- 2%*
   a) Using Python develop both an **SVM and Naïve Bayes algorithm** script to predict Class. Attach the HTML copy of your Python Code with your submission.

   Note: All steps need to be annotated (i.e. Week7-SVMNBExample)

2. **Create a PowerPoint (PPT or PPTX) presentation that includes the following:**

   a) Cover Page (Including Title, First and Last Name, Student Number)
   b) Rational Statement (summary of the problem or problems to be addressed by the PPT) – *2%*
   c) Present the Confusion Matrix/Classification Report for both the SVM and Naïve Bayes Model and Explain **three (3) key insights** from the Model Metrics (i.e., Precision, Recall, F1, Support for both summary and detailed metrics) for each model. *Note: six (6) key insights in total are required* – *6%*
   d) Compare the SVM and Naïve Bayes classification reports and present **two (2) key insights** from the Model Metrics (i.e., Precision, Recall, F1, Support for both summary and detailed metrics) with associated explanations. – *2%*
   e) Recommend one (1) model to Mr. John Hughes and present **three (3) improvements** that could enhance the usability of the model. – *3%*

Hint: Leverage the Week7-SVMNBExample

**Please post your PowerPoint Document (.ppt or .pptx) and Jupyter Notebook in HTML (.html) format via assignments under Assignment #4 by
Tuesday, November 21ˢᵗ, 2023 @ 11:59 p.m.**