Introduction to Data Analysis (DATA 1200)
Assignment #3 – Decision Tree (15% of Final Grade)
Professor: Ritwick Dutta

John Hughes wants to leverage the **raisin.csv**. He is looking at using a Decision Tree model to help predict the type of rice (Kecimen or Besni).

The dataset has 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 a **Decision Tree 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. Wk6a-DTExample)

2. **Create a PowerPoint (PPT or PPTX) 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 Correlation Heatmap and Explain **two (2) key insights** with associated explanations – *2%*
   d) Present the Confusion Matrix/Classification Report and Explain **three (3) key insights** from the Model Metrics (i.e., Precision, Recall, F1, Support for both summary and detailed metrics) - *6%*
   e) Explain **three (3) ways** to help improve the performance of the Decision Tree model. Please justify each of your answers. – *3%*

Hint: Leverage the Wk6a-DTExample

Please post your PowerPoint Document (.ppt or .pptx) and Jupyter Notebook in HTML (.html) format via assignments under Assignment #3 by

Tuesday, Nov 7th, 2023 @ 11:59 p.m.