DECISION TREES and DECISION RULES

During the next two weeks, you will develop a *multifactor* genetic test to determine if an individual may develop cancer.

This week, you will construct a two-level decision tree to classify an individual sample as a member of either the **C** class or the **NC** class. To construct the decision tree, consider the quantity “**TP** – **FP**”to select the best features for classifying samples. Complete the following steps to construct and test your tree for classifying samples:

1. Compute the quantity “**TP** – **FP**” for each feature (genetic mutation). Which feature yields the maximum value of “**TP** – **FP**”? Show the table of the top 10 features (ranked by their corresponding values of “**TP** – **FP**”) and their corresponding “**TP** – **FP**” values.
2. Select the most helpful feature, **F**, for correctly classifying samples (i.e., the genetic mutation **F** that yields the maximal value for the quantity “**TP** - **FP**”).
3. Make a confusion matrix to represent the classification of all samples with genetic mutation **F**.
4. Divide the samples into two groups by using the feature **F** to classify each sample as either
   1. group-A: samples that have mutation **F**

or

* 1. group-B: samples that do not have mutation **F**

Be prepared to provide a demo in the next class meeting, including the following:

* table of the top 10 features (see item 1 above),
* a list of samples in group-A and a list of samples in group-B (see item 3 above), and
* a confusion matrix (see item 4 above).

**NOTES:**

1. *Starting this week, you are no longer permitted to use Excel for data mining activities. You must write a computer program for all subsequent activities in the course.*
2. *You must develop your own computer program to accomplish this assignment. You ARE NOT permitted to use pre-existing programs for building decision trees or any other component of this project.*
3. *All assignments submitted for this course should represent your thinking and effort and be prepared entirely by you. Using generative AI at any stage of your work in this course constitutes academic dishonesty and violates course policy and the Ohio University Student Code of Conduct.*