# *Advanced Topics in Computer Science II (420-G50-HR)*

# *Lab 15 – Supervised Classification Models – Logistic Regression*

Date due: **April 25, 2025**

**Learning Objectives**

Upon successful completion of this lab exercise, the student will be able to:

* Use Logistic Regression for Binary Classification

Lab Set-Up

1. Download the lab folder from Moodle and rename it to *initials*G50L15 where *initials* are your initials.

To Do

**Part A – Counterfeit Finding**

1. Create a file called banknote.ipynb
2. Collect the Data
   1. Use pandas to access the csv file banknote.csv from the data\_files folder. This data file contains information on whether or not a banknote is a counterfeit based on 4 criteria measured when “imaging” the bill. The last column (authentic) is the dependent feature to be predicted.
3. Explore the Data
   1. Print the information about the dataset, the description of the dataset and the first five rows of the dataset. The dependent feature is “authentic”, all other features are independent. Check out the value set for authentic.
   2. Visualize the data (using matplotlib) and show:
      1. A bar plot of value\_counts for authentic
      2. A box plot for each of the 4 independent variables against the “authentic” feature
      3. A subplot showing notes that are authentic vs not authentic (1, 0) based on variance and skewness.
   3. For each of the plots add a comment about what the information tells you about the data (and not just it says there is a relationship…do some thinking about what it means)
4. Prepare the Data
   1. You will be using Logical Regression to build the model. From sklearn include:
      1. train\_test\_split from model\_selection;
      2. LogisticRegression from linear\_model; and
      3. Metrics
   2. Set the x and y sets for the model creation.
   3. Split the training data using the random\_state 1111.
5. Modeling
   1. Create an object of the class LogisticRegression
   2. Use the object’s fit method to create a model using the training data from above.
   3. Create a y\_pred variable from the model using the x\_test data.
   4. Check out the metrics accuracy\_score and precision\_score
   5. From metrics check out the confusion matrix of the y\_test data vs y\_pred
6. Predict More
   1. Predict using the following data:  
      -3.5637,-8.3827,12.393,-1.2823 – banknote should be authentic (1)

0.94225,5.8561,1.8762,-0.32544 – banknote is NOT authentic (0)

* 1. Use pandas to read in the file raw\_banknote.csv. This contains 250 rows of banknote data with no correct value.
  2. Run this data against your model to predict which notes are real and which are counterfeit.
  3. Compare that to the answers (matching row by row) in the file raw\_banknote\_actual.csv. How accurate was your model?

**To submit**

When you have completed the lab exercise, call the Teacher’s attention and we’ll go over it together. Then, create a single zip file called *initials*G50L15.zip and copy the file to the Moodle page for the course.