

## CS 6220 Data Mining — Assignment 8B

---

### Model Evaluation

For this assignment, we will revisit the solution you submitted to assignment 8A (the previous assignment), repeating the classification experiment that was carried out. If you used a dataset that had multiple classes, you may choose a different (binary) dataset, or you can manipulate your dataset, forcing it to be binary. To do that you can select half of your classes and simply convert them to 0, and do the same for the other half, converting them to 1.

**Objectives:** Retrain your model using the above suggestions (be sure to split your data into train/test where appropriate), and deliver a notebook containing a detailed evaluation report listing the metrics listed below, along with your commentary to each:

1. The accuracy of your model on the test data
2. The precision and recall values
3. A classification report (scikit-learn has a function that can create this for you)
4. The confusion matrix for this experiment
5. An ROC curve
6. A Precision/Recall curve

**Submission:**

Submit your file to Blackboard as before.

**Grading Criteria:**

This assignment is worth 100 points, which will be distributed according to (1) your coverage of the requirements, (2) the quality and completeness of your commentary, (3) the quality and organization of your code.