**CS5542 Big Data Apps and Analytics**

**In-class Exercise #4**

**Confusion Matrix Case Study**

We have a test dataset of 10 records with expected outcomes and a set of predictions. Calculate the Confusion Matrix for the below data(using code).

**Expected Predicted**

man woman

man man

woman woman

man man

woman man

woman woman

woman woman

man man

man woman

woman woman

Also, compute the below rates from your confusion matrix :

* **Accuracy:** Overall, how often is the classifier correct?
  + (TP+TN)/total
* **Misclassification Rate:** Overall, how often is it wrong?
  + (FP+FN)/total
  + equivalent to 1 minus Accuracy
  + also known as "Error Rate"
* **True Positive Rate:** When it's actually yes, how often does it predict yes?
  + TP/actual yes
  + also known as "Sensitivity" or "Recall"
* **False Positive Rate:** When it's actually no, how often does it predict yes?
  + FP/actual no
* **Specificity:** When it's actually no, how often does it predict no?
  + TN/actual no
  + equivalent to 1 minus False Positive Rate
* **Precision:** When it predicts yes, how often is it correct?
  + TP/predicted yes
* **Prevalence:** How often does the yes condition actually occur in our sample?
  + actual yes/total