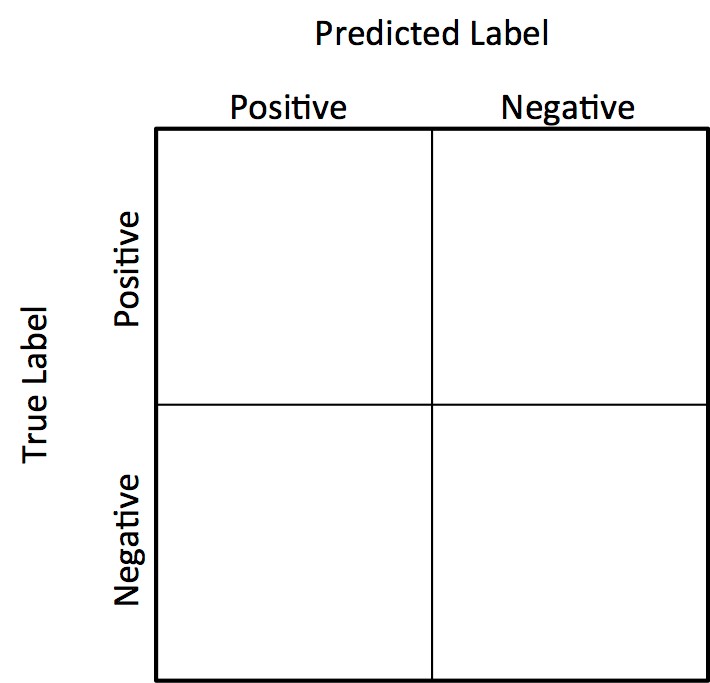
Name: \_\_\_\_Denisha Saviela\_\_\_\_\_

**CSCI 303 Introduction to Data Science Confusion Matrix Worksheet**



1. What is a confusion matrix / what can it be used for?

A summary of prediction results from classification, compares expected with predicted outcome to determine how well a model can predict unseen data (difference between y and y\_hat)

FP

False Alarm

TP

Hit

TN

Correct Reject

1. The diagonal elements are where the classifier predicted correctly; whereas, the off-diagonal elements are where the classifier made mistakes.

FN

Miss

1. Define the following terms/acronyms (and identify the one(s) we typically want to minimize):
   1. TP: true positive – correctly predicted positive examples
   2. FP: false positive – incorrectly predicted positive examples
   3. TN: true negative – correctly predicted negative examples
   4. FN: false negative – incorrectly predicted negative examples
2. Fill in the diagram with:
   1. TP/FP/TN/FN
   2. Miss/Hit/False Alarm/Correct Reject

1. The performance measures we discussed are listed below. Provide the basic definition and formula for each of:
   1. Accuracy = (TP + TN)/(TP + FP + FN +TN) – percent of correctly classified observations
   2. Precision = TP/(TP+FP) – percentage of positive prediction out of all of the examples the classifier labeled as positive what fraction were correct
   3. Recall = TP/(TP+FN) – is the percentage of positive observation as positive or out of all of the positive examples what fraction did the classifier pick up
   4. F-score = (2\*Precision\*Recall)/(Precision + Recall)– a way to balance these two, by using the F score or F measure; this is the harmonic mean between precision and recall