**Day 2:**

**Knowledge Check 1**

1. Which of the following is NOT true about the kNN classification algorithm?
   1. It is a supervised algorithm
   2. It is based on an algorithm that involves distance calculation
   3. It works with an unlabeled dataset
   4. It is primarily used for classification
2. Which of the following is relevant to cleaning data for kNN?
   1. Check for NAs
   2. Scale the predictors
   3. Make sure the target is labeled
   4. All of the above

**Knowledge Check 2**

1. Which of the following is NOT true about cross-validation?
   1. The smaller portion of the dataset is used to train the model
   2. When cross-validation is implemented, small test sets will be held out multiple times
   3. When there is not a large population on the whole, cross-validation techniques can be implemented
   4. Each subset is used as the test data set and the rest of the data is used as the training dataset
2. Which of the following functions is used to split the data into train and test sets?
   1. split()
   2. data\_split()
   3. train\_test\_split()
   4. cross\_validation()
3. Which of the following functions in Python is used to perform KNN?
   1. KNeighborsClassifier()
   2. KNNClassifier()
   3. KNearestNeighboursClassifier()
   4. None of the above
4. What is the output of the predict() function?
   1. a vector of probabilities of the predicted values
   2. the accuracy score of the classification algorithm
   3. a vector of predicted values
   4. a vector of the actual values of the target variable

**Knowledge Check 3**

1. **Select the correct formula for Accuracy**
   1. **( FP + FN ) / total**
   2. **( TP + TN ) / total**
   3. **TN / ( FP + TN )**
2. **Select the correct formula for Misclassification rate** 
   1. **( FP + FN ) / total**
   2. **( TP + TN ) / total**
   3. **TN / ( FP + TN )**
3. **Select the correct formula for True Negative Rate (Specificity)** 
   1. **( FP + FN ) / total**
   2. **( TP + TN ) / total**
   3. **TN / ( FP + TN )**
4. **[True or False] ROC is the plot of the true positive rate (TPR) against the false positive rate (FPR).** 
   1. **True**
   2. **False**
5. **[Select all that apply] Which of the following is true about the area under the curve (AUC)?:** 
   1. **The AUC should be above .5 to say the model is better than a random guess**
   2. **A perfect AUC would have a value of 0**
   3. **The AUC should be below .5 to say the model is better than a random guess**
   4. **A perfect AUC would have a value of 1**