Lab 3: Models Evaluation and Balancing

1. Load the _bank_ dataset. Train a classifier using the *Naïve Bayes* algorithms. Keep all parameters unchanged. Keep pep as the class attribute.
    a. What is the accuracy achieved?
    b. Look at the confusion matrix, and compute both sensibility (TPrate) and specificity?
    c. Look at its ROC chart.

2. Using the same dataset, train a classifier using *KNN with k=3*.
    a. What is the accuracy achieved?
    b. Look at the confusion matrix, and compute both sensibility (TPrate) and specificity?
    c. Add the new model to the previous ROC chart.
    d. How do the models compare?

3. Load the _unbalanced_ dataset. Train a classifier using the *Naïve Bayes* and *KNN (k=3)* algorithms. Keep all other parameters unchanged. Keep outcome as the class attribute.
    a. What is the accuracy achieved?
    b. Look at the confusion matrix, and compute both sensibility (TPrate) and specificity? Are these classifiers useful?
    c. Compare both classifiers through a ROC chart.
    d. Try different values for the number of neighbours (1, 10, 100). How does the global accuracy change? And their sensibility? Add them to the previous ROC chart. Are these models useful?

4. Apply different balancing techniques to the previous dataset.
    a. Use resampling increasing the number of active instances to balance the data (biasToUniformClass=1.0 and sampleSizePercent=200.0)
    b. Use SMOTE increasing the number of active instances to balance the data (percent=7000.0)

EVALUATION: How do the accuracy and sensibility change after balancing data? Are these new models more or less useful?