University of Ottawa

School of Electrical Engineering and Computer Science

CSI5155 Fall 2020

Homework Assignment 1: Bank Marketing

TOTAL MARKS 100

For this homework, please use the Portuguese Bank Marketing dataset from the UCI Machine Learning Repository. (The direct link is located at https://archive.ics.uci.edu/ml/datasets/Bank+Marketing. You should use the **bank-additional-full.csv** file with all examples (41188) and 20 inputs for this assignment.)

Instruction:

1. Submit your assignment using BrightSpace, before the due date.
2. No late assignments will be accepted.
3. This is an individual assignment.
4. Use Scikit-Learn to complete the assignment.

Question 1: Explorative Machine Learning – Binary Task [60 marks]

Import the data into your machine learning environment. Next, construct models using the following four (4) types of algorithms: a decision tree, a Naïve Bayesian learner, a support vector machine, and a k-nearest neighbor classifier. The aim of this binary learning task is to predict whether a client will purchase a product from the bank, i.e. the output variable (desired target) is feature 21, with classes 'yes' and 'no'.

Submit the following:

a. Show the four (4) models constructed by the algorithms. **[20 marks]**
b. Show the four (4) confusion matrices corresponding to the models. **[10 marks]**
c. Draw the ROC Curves to contrast the four (4) models. **[10 marks]**
d. Write a short summary, with no more than 200 words, to discuss the results you obtained and the lessons you learned when analysing this data. Your summary should include a **decision** as to which one of the four (4) machine learning algorithms you would use as well as a **motivation** for your choice. **[20 marks]**

Question 2: Multi-class learning [40 marks]

Suppose that the class label has been changed to feature 15, namely the outcome of the previous marketing campaign, with values 'failure', 'nonexistent' and 'success'. Submit the following:

a. Use a decision tree algorithm to construct a model against this multi-class problem **[10 marks]**
b. Draw ROC Curves using one-versus-one comparisons. **[10 marks]**
c. Calculate the AUC when using the one-versus-one scheme, using both the macro average and a prevalence-weighted average. **[10 marks]**
d. Write a short summary, with no more than 100 words, to contrast the results and models obtained in Questions 1 and 2. **[10 marks]**