

# MLDA Final Assignment

**Submission Deadline: Monday 5th April 2021 at 11:59 p.m.**

**NOTE:** This document is 3 pages long. Make sure you read all 3 pages as you may lose marks if you do not follow the instructions correctly.

**Task Description:** In this assignment you will develop a classifier that uses data to predict the outcome of a bank marketing campaign.

## **What you are given:**

There are two files available on Brightspace:

1. *trainingset.txt*: this file contains the training instances. It contains the descriptive features and the target feature level for each instance. You should split this dataset into training and test as demonstrated in the labs.
2. *queries.txt*: this file contains the query instances. It contains the descriptive features for each instance. However, the target feature level has been overwritten with '?'

## **Submission Deadline and Late Submissions:**

- **Deadline:** Monday 5th April 2021 at 11:59 p.m.
- Marks will be deducted for late submission - 10% per day late.

## **How do you submit your solution?**

- Please submit your solution in Brightspace under 'Assessment' -> 'Assignments' -> 'Final Assignment'.

## **What do you need to submit:**

You need to submit 2 **separate** files (**don't bundle them in a zip file, submit the two files separately**): (1) a solutions file, and (2) the jupyter (or other type of code) notebook (descriptions below).

### **1. The solutions file:**

- Naming convention: This file should be named using the following convention

'studentnumber.txt'. For example, D9876543.txt.

- **Contents:** The file should list your classifier's target variable predictions for each of the query instances in the queries.txt file. Each line in the file should list one query number followed by a comma followed by your classifier's prediction for that query, i.e.:

<query>,<prediction>

- The box below illustrates what someone looking at a portion of your solutions files should see

1,TypeA
2,TypeB
3,TypeB
4,TypeA
5,TypeA

## 2. The jupyter notebook or source code file.

- a. Naming convention: This file should be named using the following convention '**studentnumber.ipny**'. For example, D9876543.ipny.
- b. Contents: The file should contain all your code from importing libraries, data exploration, selecting features, training your model, and evaluation. Every step should be properly documented and you should reflect on choices you make as well as your experience and thoughts during the process. The notebook should be complete and I have to be able to run it on my machine and verify your implementation.

### Marking Scheme

Marks are awarded based on the accuracy of the classifier and quality of code and documentation.

### Marks may be deducted for the following reasons:

(a) Late submission (including submissions that are incomplete by the time the deadline has passed): 10 marks per day late.

(b) Incorrect submission: 10 marks will be deducted if your submission does not follow the stated formats. The reason that I do this is that if you do not correctly format your submission this slows down the correction process for everyone. I will be strict in deducting these marks. If your submission does not follow the guidelines these marks will be deducted. Examples of the types of errors that will result in these marks not be awarded include:

- Solution file named incorrectly
- Leaving blank lines between solutions in the solutions file
- Having trailing blank spaces after key values before commas in the solutions file
- Forgetting to put commas between the fields in the solutions file
- The solutions file not being a .txt file, for example submitting your solution as an .rtf or other file format (or bundling your files as a zip file submission)