

# **QBUS6180**

## **Statistical Learning and Data Mining**

### **Semester 1, 2024**

## **Classification Project: Marketing Analytics**

### **1. Overview**

In this project, your team will analyse marketing data from a bank and a retail company. Your team will have two tasks. The first will be to build machine learning models to predict the success of marketing campaigns. The second will be to uncover insights that can help your clients make better marketing decisions.

### **2. Problem description**

As a team of data scientists and business analysts working for a marketing consulting company, you have been tasked with helping two clients, a bank and a fashion store, to leverage their data to increase the effectiveness of their marketing campaigns.

The two clients provided your team with data from their latest direct marketing campaigns. You have two tasks:

1. To develop machine learning models to predict whether the marketing campaign will be successful with a customer.
2. To obtain at least three insights that can help the clients make decisions about their marketing campaigns. What types of customers are more responsive to marketing campaigns?

We will refer to these tasks as machine learning and data mining, respectively.

As part of the project, you need to write a report according to the instructions below.

### **3. Understanding the data**

#### **3.1 Two datasets**

This project involves two marketing datasets, one from a bank and another from a fashion store. The assignment requires you to work with both datasets.

One dataset primarily has numerical variables, while the other emphasises categorical variables.

### **3.2 Bank dataset**

The bank dataset is from a phone campaign to encourage clients to subscribe to a term deposit.

Each row corresponds to a call made to a customer. The response variable, *subscribed*, is the last column in the dataset. It indicates whether the client subscribed to a term deposit, which was the objective of the campaign.

The data dictionary file describes the predictor variables.

### **3.3 Fashion store dataset**

The store dataset refers to a promotional e-mail campaign.

Each row refers to a different customer. The response variable, *RESP*, indicates whether the customer responded to the promotion. It's the last column in the dataset.

The data dictionary file describes the predictor variables.

### **3.4 Data issues**

The two datasets may have issues such data errors and data leakage. Identifying and handling such issues is part of the assignment.

## **4. Machine Learning (Task 1)**

#### **Requirements for both datasets:**

- Assume a loss matrix.
- Your report must show results for at least five different sets of predictions (that is, from five different models).
- At least one of your models should be a linear model.
- At least one of your models should be a tree-based model.
- At least one of your models should be a model average or model stack.
- Identify one of your five models as a benchmark.

- Your report must compare your models in terms of cross-validation or validation metrics.
- Compare your models both in terms of the loss matrix and traditional classification metrics.
- Your report must present model evaluation results.

Note that these are only minimum requirements. Refer to the rubric for the details on the marking criteria.

## 5. Data Mining (Task 2)

**Business question:** What kinds of customers are most responsive to marketing campaigns?

**Requirements:**

- Extract at least three quantitative insights from the data that address the business question.
- You can use any combination of the two datasets for this task.

**Notes:**

- This task is open-ended, as is the nature of data mining applications. Think creatively and explore the data in a way that you find interesting. The ability to approach open-ended problems is vital in data science.
- Remember that association is not causation. Do not oversell your insights.

## 6. Written report

The purpose of the report is to describe, explain, and justify your solution to the clients. You can assume that the clients have training in business analytics. However, do not assume that they are experts on the methods used in your project.

Preparing the report will involve careful consideration of what should go in the main text. The main text should focus on the highlights of the project. Note that there is no page limit for the appendix. It's ok to put extra material (such as additional figures and tables) in the appendix and refer to it in the main text.

**Requirements:**

- Discuss problem formulation, exploratory data analysis, feature engineering, methodology, and results.

- Write about the data mining task in a separate section.
- In the problem formulation section, discuss the business problem from the perspective of decision theory. Is it a prediction problem? How can machine learning help businesses optimise their marketing efforts?
- Discuss three models in detail in the methodology section. One model should be your best linear model, the other your best nonlinear model, and the third is the model stack (or average).
- **When you submit the report on Canvas, include the Python code that generates all the results that appear in it as an additional file.**

**Suggested outline:**

1. Introduction: write a few paragraphs introducing the project and overview the methodology and main results. Use plain English and avoid technical language as much as possible in this section (write it for a broad audience).
2. Problem formulation and objectives: state the problem to be solved and the goals of the project.
3. Data understanding: provide essential information about the data, discuss potential issues, and highlight the most interesting findings. Due to a possible lack of space, you may want to refer to the appendix for most EDA plots.
4. Feature engineering.
5. Methodology: focus on the three models specified above. Explain the rationale for using these learning algorithms and explain the choices that you've made regarding configuration, training and hyperparameter optimisation. This part is allowed to be more technical than the rest of the report.
6. Results.
7. What kinds of customers are most responsive to marketing campaigns?