

**MSCI562 Coursework, Part 2**  
**Lent Term, 2022-2023**  
**Management Science Department**  
**Lancaster University Management School**

## Disclaimer

This is the second part of the two-part coursework. It assumes that you have finished the first one and thus have already done the preliminary data analysis.

## Problem Description

See the problem description in the Part 1 of the Coursework.

## Task 2. Modelling and prediction

For the different classifiers we covered in the course attempt to develop appropriate models for this problem. Your approach should predict customers satisfaction based on the available data.

- For each classification method develop one or a few candidate models that you think are promising before providing a final recommendation of the most appropriate model. Compare the approaches and select the most appropriate one. You do not need to discuss every model you tried in detail, but you must include the results for the important steps in the process that led you to the final recommendations.
- In order to ensure that the airline company makes correct decision, the marketing manager wants to make sure that:
  - 95% of dissatisfied customers are identified correctly,
  - no more than 10% of satisfied customers are mistakenly predicted to be dissatisfied.
- Explain the trade-off between correctly identifying the satisfaction vs dissatisfaction of clients. Present the trade-off visually. Explain how the results of classification would change if the goal was different?
- How many variables (and which) are actually needed to construct a sufficiently good classification model? Which of the variables can be considered as more important in the prediction of satisfaction? Explain why the selected variables are important.
- Comment on the expected generalisation performance. In other words, how well do you believe these models will perform when they are deployed?

## Guidelines

- You do not need to redo the Exploratory Data Analysis. In the text of this part, you can just refer to the part 1 of the coursework (e.g. "as we showed in Part 1 of the coursework, the variable x is not related to the variable y").
- You need to submit a report explaining your findings.
- The report should be technical, aiming at experts that understand what techniques you use. So, you do not need to explain what different techniques are.
- Your report should include the following sections:
  - Executive Summary, explaining the main findings in one paragraph. This should be in the beginning of the report;

- **Introduction**, explaining the **problems you need to address** and the **structure of your report**;
- **Conclusions**, **summarising the main findings and recommendations**.
- You can include any graphs and outputs that you find relevant to the problem in the report. Make sure that they are properly referenced and interpreted in the text. Remember that it is up to you to interpret what a figure or table shows, not for the marker to infer.
- It is also important to explain how you came to one or the other conclusion. For example, you should explain **why** you think that there is an effect in a model. Simply stating “there is” is not acceptable.
- Format sections properly, use styles (see, for example, this to learn how to do that in MS Word: <https://development.lancaster.ac.uk/mod/page/view.php?id=335>).
- Do not include any appendices (they will be discarded during the marking). You do not need to attach R code either, except for the cases, when the code helps in explaining what you did and why.
- Attach outputs of models that you are going to interpret – otherwise, how are we going to find out whether your interpretations are correct or not?
- Each student will have a different dataset, so make sure that you use the one assigned to you, not to your colleague. Failing to use the correct dataset might result in the “fail” mark for the coursework.
- Your report must be between 1000 and 2000 words and should not exceed 15 pages.

## Assessment criteria

Your work will be assessed on the following:

1. **Your ability to use correctly the tools that we covered in the module.** Your report needs to clearly show that you understand what the visualization and statistical measures you use mean, and why they are relevant to this specific problem. It is not enough to simply present relevant figures and measures; you also need to explain why you chose them.
2. **Your ability to draw correct conclusions from the visualisation and statistical measures you use.** Again including a figure or a number/statistic is not sufficient. You have to inform the reader what it means and why it is relevant.
3. **Your ability to express and justify your key findings succinctly** (rather than reporting every possible figure/table/statistic you created).
4. You will also be assessed on **report quality**, including: formatting, presenting figures and tables, having captions, numbering and proper references in the text.

You must use R in your report, all outputs should be generated using the R functions.