StirCom is a telecoms company that has spent the past few years digging up the streets of UK cities to connect people to their network. They are now keen to expand into the 5G market so they have been running a marketing campaign to encourage their existing landline customers to take out mobile contracts. Unfortunately, the call centre costs to run this campaign are high, so they would like to target customers more effectively to avoid wasted calls. This is where you come in. StirCom have provided data describing over 50000 customers and their responses to previous marketing campaigns. The data tells us if each customer responded to the marketing by setting up a new contract "new_contract_this_campaign".

The question is simple: is there any way to predict which customers are more likely to respond positively to a marketing call? Your assignment is to answer that question using machine learning techniques and produce a system that would be able to tell StirCom which customers it should target the marketing at.

You can use Orange, Python, R, or any data mining package of your choice. The data for the assignment is in a file *stircom.csv* provided for you. A data dictionary is provided describing the columns in this file.

Please note that this is an individual assignment, not group work.

<u>Requirement</u>
You should submit a report describing the modelling process you followed and your results. You should try to frame the problem in the form of CRISP-DM framework to better facilitate the discussion. Refer to the relevant CRISP-DM stages at each stage of your report. You do not need to submit code or data. The report is worth 100 marks in total and must cover the following (with weightings per section as shown):

Business Understanding [10%]
Describe the task you were given: is it clustering, classification or regression?; describe the data you received and the requirements of the finished system, including why machine learning is suitable for this task. Define any terminology that you will use in the report (for example, model, variable, task, etc.). Comment on any issues around ethics or trust that may be relevant to the framing of the problem.

Data Summary [10%]
List the variables that you found in the file provided by the company. For each one, say whether it should be treated as categorical or numeric; nominal, ordinal, continuous or discrete; and whether or not it is likely to be of use in building the solution. Explain your decisions: if you rule out any variables at this stage, you can justify your choice using summary statistics, or a histogram plot of its distribution.

Data Preparation [10%]

Describe what you did with the data prior to the modelling process. Show histograms of the data before and after any pre-processing that you carried out. (you do not need to give histograms of all variables, just the ones that need some cleaning) If you corrected any mis-typed or corrupted entries in the data, report what you changed, such as any rules you used, or examples of specific data points that were cleaned.

Modelling [30%]

You must use three different techniques and build models with each: these should include one tree-based model, one based on logistic regression, and one based on neural networks. Try to make each model perform as well as it can: if you varied the hyperparameters of a model, show which hyperparameters you varied and how this impacted on the results. Describe how you split the data for training, validation and testing purposes. Be methodical and record each result. This stage is a little like scientific research – you are carrying out experiments in your search for the best solution. Once you have a solution, show how you verified its robustness. For the three different techniques report on their comparative ability to make predictions for this problem, but only select a single model for the final test.
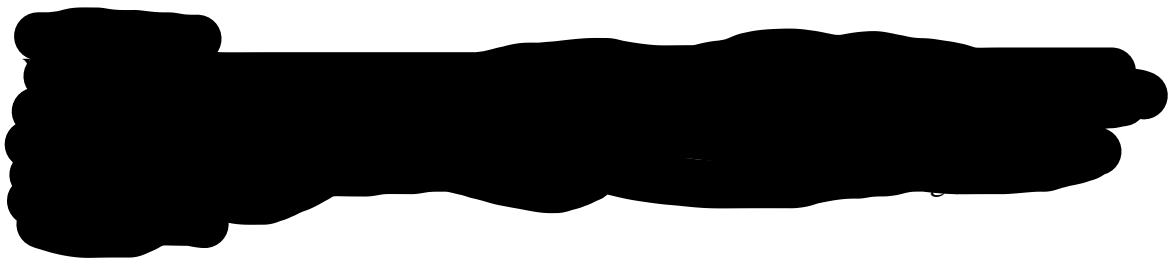
Don't try to find a perfect or extremely accurate model - one does not exist! We are interested in the procedure you followed and the justification you give for choosing particular model types/parameters/features.

Evaluation [10%]

Analyse and describe the level of accuracy the model achieves and the errors your model makes. Show a confusion matrix for your model. Are there any areas of the data where it performs worse than in others, and are there any types of error that the company would want to avoid more than others? Show a lift curve or a ROC curve for the model's predictive capability, and explain what this tells you. Comment on whether the modelling approach chosen and the results raise any trust or ethical issues.

Overall Approach and Creativity [30%]

You should adopt a structured approach to the whole process and clearly identify the five CRISP-DM stages excluding deployment in your report. Additional marks are available within this category for going beyond the basic approaches mentioned in the sections above and those covered within the course; for example, you might consider feature construction or adding some explanation to the predictions the model makes. Those showing (per the common marking scheme) originality and exceptional analytical, problem-solving and/or creative skills will be awarded the highest marks.

You ... in the "Questions" channel.

E... um word co... ll be reduc...
- From ... ve pe... nts (5 marks).
- ... the mark will be reduced by ten percent...
- W... c... old (5... if your mark was 52 and... fin...

...which is ...essment must be your own work. All students sho...
A...

Plagiarism: ...previous stud... ...GitHub or other public reposit... ...pted to in... ...ce" or "plagiarism" and will be penalized.

To avoid the risk of your ow... your... ...ng and... ...period...

...This is an individual assignment: working together with other students is not permitted. If students submit the same, or very similar work, this will be treated as "collusion" and all students involved will be penalized.

Contract cheating: Asking or paying someone else to do assignment work for you (contract cheating) is considered gross academic misconduct, and will result in termination of your studies with no award.

L... submissions: This... submiss... questions...