Analytics in Practice – Group Assignment

Transaction Prediction

Background:

You are working for an analytics consultancy company. A bank, "Universal Plus", has approached your company to request you to pitch to win a major contract with them to develop and deploy a customer transaction prediction system. This system is required to predict which customers will make a specific transaction in the future. Having such a system will ensure that Universal Plus can take proactive steps to improve their customers' satisfaction before they would take their business elsewhere. In other words, the managers' aim is to keep their promising customers. Several other consultancy companies have been approached as well, and the final decision on who will get the contract will largely depend on the outcome of a demonstration of the consultancy's approach to this problem based on a dataset that the bank provided.

The dataset[1] contains customer transactions. Unfortunately, due to confidentiality issues, the bank cannot provide the original features (variable names) and more background information about the data. There are 100,000 records and 200 anonymised numerical variables denoted by *"var_0"*, *"var_1", …, "var_199"*. Variable *"ID_code"* corresponds to the customer ID and *"target"* is the target variable and it takes value 1 if the specific transaction is made and 0 otherwise.

Presentation:

You and your team will be required to pitch your results and recommendations in a presentation of 10 minutes to Universal's "management board" impersonated by WBS staff. The pitch needs to be recorded and submitted as a video file. In the video, one should be able to see the slides (or other means of presenting) as well as the presenter – the latter is important since we also evaluate presentation skills. We leave it to you whether you want to have a single presenter or whether all of you would like to present; we will not penalise for changes of presenters (which in practice is not a good idea since it is apparently quite disruptive).

Keep it professional - business apparel is expected! Please note that this is not an exercise in professional video production; the evaluation will be based on the content of the presentation - to that end, make sure that you motivate the project properly such that it is very clear why we (as the bank) should invest in your services. Keep in mind in designing the presentation that the audience are your clients; they are managers who may not necessarily have technical knowledge so reflect carefully on what you present. They will be interested in the meaning of your work for their business, and whether their investment in this project will be worth it.

Your team also needs to submit a report in document form. Technical considerations should be described in the report, see instructions below. The report is subject to a limit of 2,000 words.

---

[1] The data actually has been obtained and adapted from Kaggle: https://www.kaggle.com/c/santander-customer-transaction-prediction/overview

Instructions:

[30 marks] Each team member should choose and review exactly one academic paper per member that is related to the data mining project. For instance, the paper can be related to the one of the machine learning techniques used in the project. Within a team, no two members should review the same paper. In your report, include a brief discussion of each of these papers. You are not expected to convey a full understanding of these papers; instead, the aim is to start working with academic literature. However, a brief justification should be given as to why you selected your chosen papers, and whether/how they may have influenced your modelling and your approach to the project.

[40 marks] Conduct data mining on the provided data to identify a suitable approach for predicting promising customers who will make a specific transaction in the future. Outline and justify your approach including data preparation, modelling and evaluation.

You do not need to provide every detailed step in the report, instead, focus on the big picture and justifications of the chosen approach and the chosen performance measures. The focus of this assignment is to give you a go at thinking about the business problem, implementing and interpreting several models using real data. A working R script should be submitted as well (of course without the data file).

[30 marks] These final marks are awarded for the sales pitch, where we will evaluate presentation style, content, media, appropriateness for the audience and generally how convincing the sales pitch is.

There is no single correct approach to this assignment; the focus is on demonstrating proper management of a realistic analytics project using the CRISP-DM methodology and that you can extract and suitably present your insights to your client.

REMEMBER THAT ALL TEAM MEMBERS MUST SUBMIT PEER ASSESSMENT MARKS BEFORE THE DEADLINE AS ADVERTISED ON MY.WBS. The peer assessment component counts 25% of the overall mark. Students who fail to submit their peer assessment forfeit the percentage of the group mark awarded by the marker equal to the impact weight of 25%. It is not possible to submit the peer assessment late.