CONTEXT

Warrior Power are an energy supplier, providing gas and electric to 5 million households across the UK. They have experienced an increase in late and defaulted payment over the last twelve months and are seeking to understand any reasons for this. In particular, understand which customers are likely to experience difficulty paying bills so that preventative measures can be taken.

They have provided CACI with a representative sample of electricity customers, and asked us to provide some initial analysis – the aim being to scale up and roll out a predictive model onto the entire customer base.

TASK

Please find attached the Warrior Power customer data. This has been matched and coded with CACI “Ocean” data, which will provide additional features for any modelling algorithms. There is a brief data dictionary provided.

Please provide a Python notebook(s) that go some way to meet Warrior’s objectives. That is, recommend a predictive model that aims to identify or classify electricity customers that are most likely to have difficulty paying bills. You should:

* Present the accuracy/ robustness of your model(s)
* Share and discuss the importance and validity of the features/ classifiers within your model(s)

You should assume that the audience for this part of the assessment is a technical one, such as for an internal peer review. Following this we may ask you to describe or explain elements of your work to a non-technical audience, such as part of client delivery to Head of Strategy, CMO, or similar.

GUIDELINES

A good solution to this exercise will:

* Be clear and concise; with well-designed and efficient code, commented as appropriate; effectively presented, avoiding unnecessary jargon
* Recommend a robust model that can be easily applied to the Warrior customer base to produce an indication of default risk
* Consist of a logical and easily-understood algorithm, which can be easily sold into the client’s organisation

Note, the overall accuracy of your model(s) is not being assessed – the process and interpretation are the important elements.

Your model(s) should be presented and discussed as part of the interview process, where no more than 20 minutes has been allocated. Therefore please be concise, and do not produce more output than can be reasonably covered!

Finally, we do not expect you to commit unreasonable amounts of time to this exercise. Most candidates spend no more than 1-2 hours completing this task.