Experian Workshops 2021

Development and Quality Assurance

Experian's wide client base of banks, lenders, insurance companies need a software for assessing the credit eligibility of their customers. The tool (product, called MyCS) needs to calculate the credit score based on an input set of data where each record represents an entity for the credit score check. The result of the check should be provided in a form that is usable for both manual check and automated data processing.

The data can be loaded both manually by a credit check operator (employee) or by a system on the file system of MyCS.

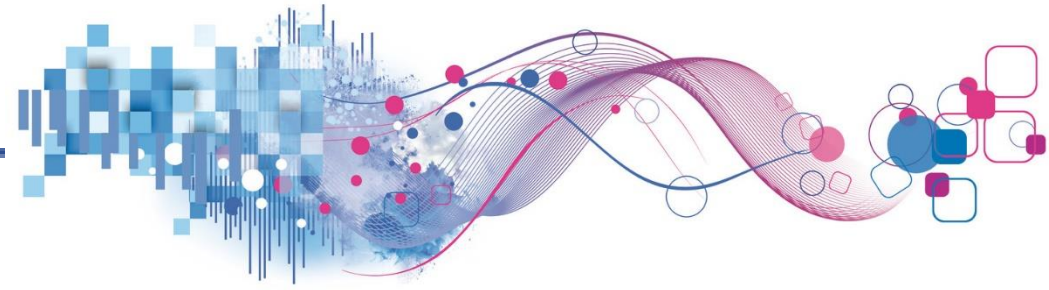
MyCS should:

- **be able to process CSV data initially assuming that all clients provide the data in one and the same structure**
- **expose an API for file submission for batch processing – load files into the scoring model**
- **expose an API for manual checks of single records**
- **provide valuable information in case of wrong data structure or invalid data**
- **abort the processing of data in case of processing failure**
- **log on the file system all events of the processing**
- **not amend or delete the data, the structure, or the file with or without data processing or failure**
- **should provide the result as a report on the file system**
- **not process a file more than once**
- **UI is nice to have and should be able to cover each API endpoint**

The software must be tested from back-end and front-end perspective. Quality Assurance must ensure software is functioning as desired and as many test case scenarios are covered.

Analytics

MyCS very nature will be a statistical algorithm that assess customer eligibility for credit products. You need to create a model using Linear Regression with dummy variables. The model should differentiate between Good and Bad customers, i.e. customers who are likely to pay their loan and those who are less likely to do so.



At minimum:

- **There must be no inversions in the coefficients of the dummy variables**
- **The model must validate when using the Kolmogorov–Smirnov test**

A good model would also meet these requirements:

- **It will contain between 7 and 15 variables**
- **There is a good score distribution, without large clusters of the population receiving the same or very similar score (e.g. 5% of the population all receiving 274 points)**
- **The Gini coefficient of the model is as high as possible. For an application model, like the one produced here, you should aim for a Gini coefficient of at least 50-60 points.**
- **Create documentation supporting model explainability.**