Capstone 2 – Project Proposal
Credit card approval prediction model using vintage analysis

Introduction

Vintage analysis is a method used for managing credit risk. It illustrates the behavior after an account was opened and identifies if accounts opened are riskier than others. Data is grouped into segments based on the original month. The model measures the risk of transactions by identifying probability of default/risk over a specific amount of time. The risk has been flagged by a binary indicator provided in the dataset.

Dataset

- https://www.kaggle.com/rikdifos/credit-card-approval-prediction
- 2 tables of data merged by ID (application record and credit record)
- Data has no prepared labels and must be constructed
- Rows = months of opening accounts and columns = months after opening
- Values are accumulate past-due rate
- 439,000 applicants

Approach

- Will use vintage analysis to construct label
- Define ‘Good’ and ‘Bad’
- Observe window needs to be wide and we analyse the behaviour of customers prior to the applications being submitted and aggregate features as of application
- We will analyse customer behaviour and identify traits/patterns by Exploratory Data Analysis
- We will also run statistical tests against each hypothesis in the data and use the results to create additional variables

- We will test the variability and explanatory capacity of the default and engineered features by using them in a model that will quantify the likelihood of Bad behaviour of the customer/applicant
- We will measure the performance of the model and validate the results by running diagnostics and quantify overfit of the model
- Further steps will be documented based on the maturity and opportunity areas of the current solution

Artefacts to be submitted

- Data wrangling notebook
- Exploratory Analysis notebook
- Statistical Analysis notebook
- Data Story document
- Model development notebook
- Milestone report
- Capstone presentation
- Additional excel files/Analysis as a part of the capstone