Final Project Proposal – Group 3

- N.Ashraff – s14308
- D.S.Sonnadara – s14508
- D.M.P. Disanayaka – s14472

Description of the problem

Credit cards are a commonly used mode of transaction in the modern world. With the number of users constantly rising, banks need to streamline their application procedures to ensure optimal efficiency in choosing customers to offer their services to. The aim of this analysis is to create a product that will be able to accurately filter good prospective customers from the bad.

Description of the dataset

Feature name	Explanation	Remarks
ID	Client number	
CODE_GENDER	Gender	
FLAG_OWN_CAR	Is there a car	
FLAG_OWN_REALTY	Is there a property	
CNT_CHILDREN	Number of children	
AMT_INCOME_TOTAL	Annual income	
NAME_INCOME_TYPE	Income category	
NAME_EDUCATION_TYPE	Education level	
NAME_FAMILY_STATUS	Marital status	
NAME_HOUSING_TYPE	Way of living	
DAYS_BIRTH	Birthday	Count backwards from current day (0), -1 means yesterday
DAYS_EMPLOYED	Start date of employment	Count backwards from current day(0). If positive, it means the person currently unemployed.
FLAG_MOBIL	Is there a mobile phone	
FLAG_WORK_PHONE	Is there a work phone	
FLAG_PHONE	Is there a phone	
FLAG_EMAIL	Is there an email	
OCCUPATION_TYPE	Occupation	
CNT_FAM_MEMBERS	Family size	
MONTHS_BALANCE	Record month	The month of the extracted data is the starting point, backwards, 0 is the current month, -1 is the previous month, and so on
STATUS	Status	0: 1-29 days past due 1: 30-59 days past due 2: 60-89 days overdue 3: 90-119 days overdue 4: 120-149 days overdue 5: Overdue or bad debts, write-offs for more than 150 days C: paid off that month X: No loan for the month

There are 438557 observations and 20 variables in the dataset

Problems and concerns

The response variable in this dataset does not predict whether customers are good or bad, but rather gives them a credit score. Thus we need to find a method to classify a prospective customer as good or bad using the status of their credit score.

Click here to access the dataset