Credit Card Approval Prediction

Abstract:

Banks receive a lot of applications for issuance of credit cards. Many of them are rejected for many reasons, like high-loan balances, low-income levels, or too many inquiries on an individual's credit report. Manually analyzing these applications is error-prone and a time consuming process. This task can be automated with the power of machine learning. In this project, we will build an automatic credit card approval predictor using machine learning techniques, just like the real banks do.

Problem Statement:

The task is to analyze and build a predictive model that can accurately determine whether a credit card application should be approved or rejected based on various applicant attributes and historical credit data. The goal is to develop a model that can assist financial institutions in making informed decisions while minimizing the risk of default and maximizing profitability.

They have given a problem to identify the customers segments which are eligible for Credit Card approval, so that they can specifically target these customers.

Variable Description:

File - Application Record.csv

Column	Description
ID	Unique Id of the row
CODE_GENDER	Gender of the applicant. M is male and F is female.
FLAG_OWN_CAR	Is an applicant with a car? Y is Yes and N is NO.
FLAG_OWN_REALTY	Is an applicant with realty? Y is Yes and N is No.
CNT_CHILDREN	Count of children.
AMT_INCOME_TOTAL	the amount of the income.
NAME_INCOME_TYPE	The type of income (5 types in total).
NAME_EDUCATION_TYPE	The type of education (5 types in total).
NAME_FAMILY_STATUS	The type of family status (6 types in total).

DAYS_BIRTH	The number of the days from birth (Negative values).
DAYS_EMPLOYED	The number of the days from employment (Negative values). This column has error values.

FLAG_MOBIL	Is an applicant with a mobile? 1 is True and 0 is False
FLAG_WORK_PHONE	Is an applicant with a work phone? 1 is True and 0 is False.
FLAG_PHONE	Is an applicant with a phone? 1 is True and 0 is False.
FLAG_EMAIL	Is an applicant with an email? 1 is True and 0 is False.
OCCUPATION_TYPE	The type of occupation (19 types in total). This column has missing values.
CNT_FAM_MEMBERS	The count of family members.

File - Credit Record.csv

ID	Unique Id of the row in application record.
MONTHS_BALANCE	The number of months from record time.
STATUS	Credit status for this month. X: No loan for the month C: paid off that month 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

Note –

- DAYS_BIRTH ---> Count backwards from current day (0), -1 means yesterday.
- DAYS_EMPLOYED ---> Count backwards from current day (0). If positive, it means the person currently unemployed.
- MONTHS_BALANCE ---> 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 ---> 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, writeoffs for more than 150 days C: paid off that month X: No loan for the month

Scope:

- Understand data by performing exploratory data analysis
- Training and building classification algorithm to predict if a customer will be approved with credit card or not
- Understand feature importance and improve the model
- Understand various model performance metrics and measure the performance of each model

Objective & Learning Outcome:

The objective is to train a machine learning model using the provided dataset to predict the approval outcome for new credit card applications accurately. The model should generalize well to unseen data and provide a reliable assessment of the creditworthiness of applicants.

Students should be able to predict credit card approval from records with the help of a classification model. They should also be able to perform EDA and re-build the model and check if there is any significant change in the predictive scores.