Credit Card Defaulters

You are given dataset of credit card holders to identify them as defaulters/non-defaulters (loan paid or not) based on their respective attributes.

Dataset Description

Column Name	Column Description
TARGET	Target variable (1 – defaulter, 0 – non-defaulter)
CONTRACT_TYPE	Identification if loan is cash or revolving
GENDER	Gender (M/F)
FLAG_OWN_CAR	Flag if the client owns a car
FLAG_OWN_REALTY	Flag if client owns a house or flat
CNT_CHILDREN	Number of children the client has
INCOME_TOTAL	Income of the client
CREDIT	Credit amount of the loan
INCOME_TYPE	Clients income type (businessman, working, maternity leave...)
EDUCATION_TYPE	Level of highest education the client achieved
FAMILY_STATUS	Family status of the client
HOUSING_TYPE	What is the housing situation of the client (renting, living with parents, ...)
OCCUPATION_TYPE	What kind of occupation does the client have
FAM_MEMBERS	How many family members does the client have
REGION_RATING_CLIENT	Rating of the region where client lives (1,2,3)
ORGANIZATION_TYPE	Type of organization where client works

Deliverable

The project will have three deliverables. Keep on adding the deliverables to the same document. This way, the final report will be a comprehensive report.

1. Exploratory Data Analysis (EDA) – 40% weightage

2. Clustering and Frequent Pattern Mining -40% weightage

3. Recommendations Based on Insights from Data/Final Report -20% weightage

1. Exploratory Data Analysis (EDA)

This deliverable will be focused towards getting familiar with the data. For that you will have to get your hands dirty with the data. This will consist of initial data exploration, visualizations etc. This could include but not limited to the following points:

- Data Preprocessing (Identifying duplicate data and removing it, finding missing values and replacing them by an appropriate strategy, normalizing the data)
- Calculating and visualization of summary statistics (consider boxplots, bar graphs)
- Finding correlation between attributes.
- Finding out dependence between categorical and numerical attributes.
- Using graphs to visualize the data (at least 5-6 graphs representing meaningful information)

2. Clustering and Frequent Pattern Mining

This deliverable will be more focused toward features selection and engineering for data mining tasks. Based on a selected set of features you will perform specific tasks that can extract useful insights from data.

We are just listing directions to think, you can come up with something else by which more meaningful information from data can be deduced.

- Perform frequent pattern mining to identify patterns of defaulters and non-defaulters.
- Perform outlier analysis on the data.
- Perform cluster analysis on the data.
- Report your findings.

3. Final Report

For this deliverable, focus on the following points.

Try to apply all the concepts learned during the course. You should be able to identify the approaches and techniques. You need to explain what information you have extracted from the data and suggest ways to help the bank department reduce the number of defaulters. Your conclusions, recommendations, reasoning, and findings should be supported by the relevant graphs/visualizations and figures.

Dataset Link:

https://drive.google.com/file/d/1ULbDRmzmy0C-aHYGxmM5_a-iM8DlbJ-w/view?usp=sharing