Domain

In the banking and financial domain, assessing the creditworthiness of an applicant is crucial. Banks receive a plethora of credit card applications and require a proficient method to predict the approval or rejection of these applications while minimizing the risk of default. This task would help banks and other financial institutions streamline their application processes, make improved decisions, and reduce the risk of approving applicants likely to default on their payments.

The dataset for this project, sourced from Kaggle, is derived from actual credit card applications at a bank. Provides a practical context for applying machine learning techniques to solve a real-world problem.

Project Overview

The primary goal of this project is to build a model that accurately predicts whether a credit card application will be approved or not. The model will be trained on a dataset of credit card applications that includes various features such as income, employment status, and family status, among others.

This is a binary classification problem, where the outcome can be one of two possible options: 'approved' or 'not approved'. The predictive model will be evaluated based on its accuracy, precision, recall, and F1 score. Further, the project will explore the relationships and trends between the various features and the application outcomes.

In addition to creating a predictive model, this project will provide a comprehensive analysis of the dataset, including exploratory data analysis (EDA) and feature importance identification. The results of this analysis will provide valuable insights into the demographics and other features that influence the approval of credit card applications.

Learning Outcome

Upon completion of this project, you will:

- Gain an understanding of the domain context
- Learn to preprocess and clean data to prepare it for analysis
- Gain experience in conducting exploratory data analysis to identify patterns and trends
- Build and evaluate predictive models using machine learning algorithms
- Learn to interpret the results and make recommendations

Data Description

The dataset comprises 5,37,677 credit card application records with 19 features. Each record represents an individual application.

Attribute Information

The dataset attributes include:

S.no	Feature Name	Description	Remarks
1	ID	Client Number	Unique Identifier for each client
2	CODE_GENDER	Gender	Gender of the client
3	FLAG_OWN_CAR	Is there a car	Whether the client owns a car
4	FLAG_OWN_REALTY	Is there a property	Whether the client owns property
5	CNT_CHILDREN	Number of Children	Number of children the client has
6	AMT_INCOME_TOTAL	Annual Income	Total income of the client
7	NAME_EDUCATION_TYPE	Education Level	Level of highest education the client achieved
8	NAME_FAMILY_STATUS	Marital Status	Family status of the client
9	NAME_HOUSING_TYPE	Way of Living	What is the housing situation of the client
10	DAYS_BIRTH	Age in days	Client's age in days at the time of application
11	DAYS_EMPLOYED	Duration of work in days	How many days before the application the person started their current employment
12	FLAG_MOBIL	Is there a mobile phone	Does the client have a mobile phone
13	FLAG_WORK_PHONE	Is there a work phone	Does the client have a work phone
14	FLAG_PHONE	Is there a phone	Does the client have a phone

15	FLAG_EMAIL	Is there an email	Does the client have an email
16	JOB	Job	The type of job the client has
17	BEGIN_MONTHS	Record month	The month of the extracted data is the starting point, backward, 0 is the current month, -1 is the previous month, and so on
18	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
19	TARGET	Target	Risk users are marked as '1', else are '0'

Questions

1. What is the distribution of approved and not approved applications in the dataset?
2. Which attributes have the most influence on credit card approval?
3. Can we establish any relationships or trends between the features and the application outcomes?
4. What is the demographic profile (age, gender, income level, etc.) of applicants with high approval rates?
5. How does the number of dependents influence the approval of the application?
6. How do the employment status and job type of an applicant affect the approval decision?

References

Lao Tse. (2020, March 10). Credit Card Approval - With Target.