

Project Proposal: Credit Card Approval Prediction using Machine Learning

Introduction:

Credit card approval is a crucial process that involves evaluating an individual's financial history, creditworthiness, and other factors. Banks and other financial institutions must make informed decisions when approving credit card applications to minimize risks and maximize profits. In this project proposal, we aim to build a machine learning model that can accurately predict credit card approval based on various features of the applicants.

Dataset Description:

The dataset used in this project is provided by Kaggle and contains information on credit card applications from a bank. The dataset comprises 4,00,000 observations with 20 explanatory variables, including numerical, categorical, and binary variables. The target variable is a binary variable that indicates whether the credit card application was approved or not. The dataset is split into two files – application_record.csv and credit_record.csv.

Column Name	Column Description
ID	Client number
CODE_GENDER	Gender
FLAG_OWN_CAR	Does the client own a car
FLAG_OWN_REALTY	Does the client have a property
CNT_CHILDREN	Number of children in the family
AMT_INCOME_TOTAL	Annual Income
NAME_INCOME_TYPE	Income category
NAME_EDUCATION_TYPE	Educational Level
NAME_FAMILY_STATUS	Marital Status
NAME_HOUSING_TYPE	Way of Living
DAYS_BIRTH	Date of Birth
DAYS_EMPLOYED	Start date of employment
FLAG_MOBIL	Does the Client have a mobile phone
FLAG_WORK_PHONE	Does the Client have a work phone
FLAG_PHONE	Does the Client have phone
FLAG_EMAIL	Does the Client have email
OCCUPATION_TYPE	Occupation Type
CNT_FAM_MEMBERS	Size of Family
ID	Client number
MONTHS_BALANCE	Record month
STATUS	Status

Project Objective:

This project aims to build a machine learning model that can accurately predict the credit card approval status of an applicant based on various features such as income, education, age, and credit history.

To achieve this objective, we will follow the following steps:

1. **Data Preprocessing:** In this step, we will examine the dataset for missing values, outliers, and other issues affecting the model's accuracy. We will then impute missing values using appropriate methods such as mean imputation or regression imputation. We will also encode categorical variables and perform feature scaling to prepare the data for modeling.
2. **Exploratory Data Analysis:** In this step, we will perform various statistical analyses and visualizations to understand the relationships between the variables and identify any patterns and trends in the data. We will examine the distribution of the target variable (credit card approval status) and look for any correlations between the target variable and the explanatory variables. We will also create scatter plots, box plots, and histograms to visualize the distribution and relationships between the variables.
3. **Feature Selection:** In this step, we will use feature selection techniques such as correlation analysis, mutual information, and recursive feature elimination to identify the most critical variables that contribute to the prediction of credit card approval. We will then select a subset of the most important variables for use in the classification models.
4. **Model Selection:** In this step, we will experiment with classification models such as **logistic regression, decision tree, random forest, support vector machines** and **XG boost** to select the best model for predicting the credit card approval status. We will train the models on the training dataset and evaluate their performance using cross-validation. We will also tune the hyperparameters of each model to optimize its performance.
5. **Model Evaluation:** In this step, we will evaluate the performance of the model using various metrics such as accuracy, precision, recall, F1 score, and area under the receiver operating characteristic (ROC) curve. We will also create visualizations such as confusion matrices to identify any areas for improvement in the model.

Overall, this project aims to develop a robust and accurate machine learning model that can predict the credit card approval status of an applicant based on various features. The model will have applications in the banking and financial industry for approving credit card applications and minimizing risks associated with credit card fraud and default.

Conclusion:

This project results can be used by banks and other financial institutions to make informed decisions regarding credit card approval. Additionally, the project can be extended to other financial institutions and countries to develop similar models for predicting credit card approval status.

Link to Dataset:

<https://www.kaggle.com/datasets/rikdifos/credit-card-approval-prediction>

Purpose of choosing the dataset:

During an interview for a finance-based organization, I was asked to summarize the credit card approval ML model, but I was unable to provide a detailed response. This experience made me realize that I lacked the necessary knowledge and skills to explain the model effectively. Therefore, I would like to work on the credit card approval ML model as part of my project.

My purpose for choosing this dataset is to gain a deeper understanding of the model and its implementation. I believe that working on this project will help me improve my knowledge of the underlying data and techniques used to build the model. By completing this project, I hope to develop a stronger foundation in machine learning and finance, which will be useful for future career opportunities.