Credit Card Approval - Classification Model

Final Project Proposal
Machine Learning 1 - BIA 5320
Professor Bushra Bashir

Group 5

Banshi Keshwala
Sam Leslie
Ishita Vaghela
Labdhi Zatakia

Problem description:

Financial institutions face risk when it comes to issuing credit, since some customers may default on their payments due and cause debt to accumulate for the institution. Providing credit as a service is the highest source of profitability that a financial institution has, and since that profitability comes from resolving balances, there is a major risk factor associated with customers who cannot resolve their debts.

These institutions do not want to face losses associated with bad customers, so finding accurate classification and prediction models is very valuable to these companies, as they can be a determining factor in overall profitability. The goal of these models could be to determine which potential customers to approve, what their interest rate should be, and what their minimum payment amount should be.

Typically these models will analyze existing customer data related to their personal finances and financial behavior. This could include assets owned, gross income, marital status and past behaviors when it comes to loan payments. Using this data, it is possible to create a model which can predict outcomes if a particular person would be faulty on their loan payments, and therefore approved or not approved for a credit card. Based on this model, we can classify future potential customers to allow for the greatest possible outcome for a financial institution in terms of risk management.

Our classification model will seek to rank customers based on their existing data and give them a good, medium or bad ranking, and in addition it will also give a “probability of default” rating to reflect grey areas in the outcomes of the model. This will allow the decision to rest on the financial institutions based on their best interest considering the current climate of their firm.

Machine learning is relevant to solving this problem since the dataset is large, and analyzing it through manual queries and basic visualizations will not be sufficient to create an accurate probability guess for each potential customer. Applying the automated benefits of machine learning to this problem will allow for efficient result processing, and a cleaner output model than can be used repeatedly by its end user to generate outputs.

Dataset Details:

Source: Kaggle - Dataset

Type of data: Table

Size of dataset: 438,557 rows and 18 columns in application record, 1,048,575 rows and 3 columns in credit record

Preprocessing steps: Combining both datasets into one and checking for duplicate entries and deleting duplicates, checking for any blank values and accordingly filling spaces, looking for outliers in the dataset and correcting it as well as balancing the dataset.

Dataset description:
Application_record - In application record there are details about customers' personal information and we can use that to build a model to calculate probability for the applicant's default for payments. This dataset consists of 18 columns and has information about marital status, number of children, number of properties and cars, etc. and many other information.

Credit_record - This data set contains only 3 columns; ID, Months_balance, and status. It informs of the status of balance payments for the customer, and indicates which month the entry is referring to. It is crucial to understanding a customer's timeliness in paying their credit loan.