

CREDIT RISK ASSESSMENT

PROJECT EXECUTIVE SUMMARY

Goal and Introduction:

My goal was to develop a reliable method for classifying credit risk using a dataset comprising 100,000 customer records with 27 characteristics. I meticulously analyzed the dataset to identify key variables, made necessary adjustments, and cleaned it up. By averaging monthly records from January through August into a single representative column, I streamlined the dataset to reflect consolidated customer information. I employed various models, such as KNN, decision trees, and model fitting, to evaluate their accuracy, achieving an overall accuracy rate of 83.11%.

Preprocessing Steps:

I modified the dataset to enhance analysis efficiency and insight. Specifically, I restructured the data by creating eight new columns for each variable, corresponding to each month's entries. This change enabled me to reduce the data volume per customer to a more manageable size. For instance, for the variable 'Annual Income', I generated eight columns, each representing a month's data. When monthly entries for a given variable were consistent, I retained only one column and removed the unnecessary seven. I then calculated the mean of these columns for every variable to eliminate duplication and extract significant insights. I meticulously addressed outliers and missing entries, implementing necessary modifications to ensure data integrity. Using this refined dataset, I tested my initial hypothesis regarding the relationship between credit score and customer risk level through several model-based evaluations.

Business Question:

How can I more effectively identify high-risk and low-risk customers to make informed lending decisions while reducing manual effort?

Business Model:

The primary aim of my business model was to efficiently categorize customers into "Low-risk" and "High-risk" groups based on their credit-related data. The target variable, `credit_score`, initially included three categories: 'Good,' 'Bad,' and 'Standard.' I reclassified "Good" and "Standard" as "Low risk" due to favorable credit ratings, and "Bad" as "High risk" due to adverse credit conditions. My models significantly enhanced operational efficiency by automating the credit risk assessment process, offering quick, reliable, and consistent results. This streamlined approach helps financial institutions make informed decisions, effectively manage risks, and optimize resource allocation.

Best Model:

The decision tree emerged as the best-performing model in my dataset.

Business Decision:

Implementing my credit risk classification models enables banks to make more prudent loan approvals. By classifying customers as "Low risk" or "High risk," financial institutions can decide the feasibility of granting loans. For example, my findings might prompt banks to reconsider lending to high-risk clients, leading to more judicious resource use and a reduction in potential financial risks. This strategic decision-making process aligns with my ultimate objective of improving the overall efficacy and risk management procedures of financial institutions.