Credit Risk Analysis Report

The purpose of this analysis is to predict the risk of loans based on historical lending to build a model that can identify the creditworthiness of the borrowers.

The dataset consists of loan size, interest rate, borrower income, debt to income, number of accounts, derogatory marks, and total debt.

· The data set was split into Training and Testing sets

· The loan status column indicates if a loan is either a healthy loan (0 value) or has a high risk of defaulting (1 value)

· A Logistic Regression Model was created with the original data to predict both health and high-risk loans (see Model 1 results below)

· The training data was then resampled using the RandomOverSampler module and that data was used to build a new Logistic Regression Model. The resampled data was then used to predict both the healthy and high-risk loans (see Model 2 results below)

Results

Machine Learning Model 1:

· Accuracy is 99%

· Precision for Healthy Loan is 1.00, High-risk loan is .87

· Recall for Healthy Loan is 1.00, High-risk loan is .89

Machine Learning Model 2:

· Accuracy is 100%

· Precision for Healthy Loan is 1.00, High-risk loan is .87

· Recall for Healthy Loan is 1.00, High-risk loan is 1.00

Summary

· Logistic regression model with the oversampled data was better at making predictions

·The regression model with the oversampled data would be better to use in this case because it is more accurate at predicting the high-risk loans, and those are the types