Credit risk model for forecasting loan default

W.K.T.H.Madumal(SC/2019/10737), N.G.S.S.Nanayakkara(SC/2019/10799),
B.K.V.Jayarathna(SC/2019/10914)

INTRODUCTION

Our Credit Risk Model is an advanced forecasting tool that carefully examines a range of financial data to anticipate loan failure. This model evaluates borrowers' credit worthiness using statistical algorithms, offering crucial insights for risk management. Our methodology minimizes the impact of prospective defaults by enabling financial institutions to make well-informed lending decisions by integrating historical data, market trends, and machine learning. An effective and versatile instrument that improves loan portfolio management, protects assets, and eventually fortifies your institution's financial stability.

METHODOLOGY

The first step involves cleaning the data to ensure its quality and consistency. This process includes removing duplicate records, correcting formatting errors, standardizing variables, and addressing outliers. We will carefully evaluate each variable's relevance and usefulness in the analysis. Variables that do not contribute significantly to the prediction of credit risk will be excluded from further analysis. The association between each independent variable and the dependent variable (credit risk) is taken into consider while selecting covariates using covariances. Due to their substantial positive relationship with credit risk, variables with a high positive covariance with the dependent variable is more likely to be included as covariates.

Now build a model for the risk of loan default using binomial regression. Selecting among three link functions: logit, probit, cloglog, with justification for final selection. Testing and validating the assumptions of the model using Durbin-Watson, scale location plots.

Packages:

openxlsx, ggcorrplot, randomForest, caret, glm, dplyr, ggplot2, AER, lmtest

Assumptions:

Independence of observations, Homoscedasticity, Deviance Residual, Overdispersion

RESULTS

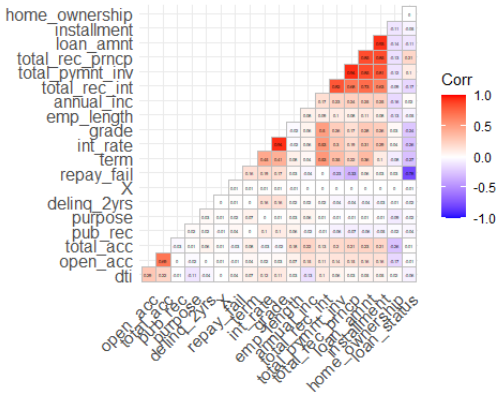

Figure: The covariance matrix between independent variables.

Model	Link function	AIC Value
Model 01	Logit	1530.4
Model 02	Probit	1602.3
Model 03	Cloglog	1615.1

Table: AIC values of link functions

CONCLUSIONS

Comparing the AIC values minimum value obtain Logit function. Therefore, the suitable model is model 01.

Generalized linear modelling is best method to analysis of real-world customer loan data. Overall, this analysis contributes to a deeper understanding of credit risk factors and lays the groundwork for informed decision-making in the realm of loan forecasting.