

AY2022 BC2406 Computer Based Assessment Homework

Home Loan Approvals

Introduction

Providing Home Loans is a relatively safe and profitable business for many banks and financial institutions as the property can be used as collateral. The CEO of a bank is thinking about using Analytics for the loan approval process.

To study the feasibility, a sample of home loan data (homeloan2.csv) was handled to you. The definitions of the data variables are provided in Appendix A.

1. Data Checks:
 - a. Check and correct (if necessary) the data type for each variable in the dataset. List the datatype (integer, factor, ...etc) for each variable.
 - b. List the number of missing values for each variable. How would you handle missing values in the dataset? Explain. Execute data cleaning (if any).
2. Explore the data using tables, statistics or charts and report on your key findings. Justify and execute data cleaning (if any). How many cases are in your final cleaned dataset?
3. Using CART and one other model (in this course) on your final cleaned dataset:
 - a. Should Loan_ID be used as a predictor X variable?
 - b. Evaluate the predictive accuracy and state the best model found. Produce a table that shows the predictive accuracy of the 2 models.
 - c. Identify the key factors that determines Loan Status.
 - d. What type of prediction error is more serious?
4. Suggest a way to further reduce the more serious prediction error [No coding].
5. Is there evidence of gender discrimination in loan approved? Explain. State the dataset used for this analysis (original dataset as given, final cleaned dataset, trainset, testset, ...etc).
6. What can be done to improve the success of this analytics in the bank?
[Maximum 1 page].

Variable Name	Description
Loan_ID	Unique Loan ID
Gender	Male/ Female
Married	Applicant married (Y/N)
Dependents	Number of dependents
Education	Applicant Education (Graduate or Not a Graduate)
Self_Employed	Self employed (Y/N)
ApplicantIncome	Applicant income (monthly average)
CoapplicantIncome	Coapplicant income (monthly average)
LoanAmount	Loan amount in thousands
Loan_Amount_Term	Term of loan in months
Credit_Score	0: Bad Credit Score; 1: Good Credit Score.
Property_Area	Area Location A, B or C.
Loan_Status	Loan approved (Y/N)