Prediction of Loan Risk

Introduction

Segmenting data based on demographic factors is a widely used practice in marketing including the banking sector despite the fact that the correlation of these factors where the customers are concerned is often weak. Risk assessment is an important task of bank, as the increase and decrease of credit limits in bank depends largely to evaluate the risk properly. The key problem consists of identifying good and bad customer's status those who applied for loan. In order to increase the accuracy of risk, risk assessment is performed in primary and secondary levels. The objective of the research project is to find assess the risk to determine whether the loan can be sanctioned to a customer or not.

Bank wants to automate the loan eligibility process based on the customer details provided while filling the online application form. To do so bank segments such collected customer data based on expected benefits and attitudes to determine customers who are eligible for loan amount so that they can specifically target those customers. This provides the bank with the advantage to address the conflict/gap between individual services and cost-centric/cost-saving standardization.

About the Dataset

The dataset for the project consists of a set of customers of a bank with different attitude-dimensions. The goal is to find whether the customers were eligible to get a loan sanctioned or not.

Variable	Description
Loan_ID	Unique Loan ID
Gender	Male/ Female
Married	Applicant married (Y/N)
Dependents	Number of dependents
Education	Applicant Education (Graduate/ Under Graduate)
Self_Employed	Self employed (Y/N)
ApplicantIncome	Applicant income

Variable	Description
CoapplicantIncome	Coapplicant income
LoanAmount	Loan amount in thousands
Loan_Amount_Term	Term of loan in months
Credit_History	credit history meets guidelines
Property_Area	Urban/ Semi Urban/ Rural
Loan_Status	Loan approved (Y/N)

Challenge

1. Perform Exploratory Data Analysis to understand the significant features that determine the loan status.
2. Develop a K-Means Clustering model to segment the customers who are eligible for loan amount.