

Capstone 2 Proposal

Loan Default Prediction

Problem: Banks are faced with the issue of loans not being paid back. People who can't pay back the loan due to bad financial management, losing their jobs or who took the loan in a stressful situation may suddenly not be able to make the loan payments.

Instead of suddenly finding out that these people cannot pay their loans, this project aims to predict these high risk loans and take pre-emptive actions.

Statement: Predicting currently active loans that are most likely to default based on loanee information and his/her financial status.

Client: Banks, loan issuing facilities.

By solving this problem, banks can:

1. Prepare a budget for predicted loans that will default
2. Take pre-emptive measures against loans that are most likely to default.
3. Prepare more subjective criteria when approving loans.

Dataset:

This dataset was found on Kaggle and was originally bought from Lending Club Loan Data.

<https://www.kaggle.com/wendykan/lending-club-loan-data>

Outline:

This problem is best solved using supervised classification methods as we have data from loan data along with their status from 2007 – 2015.

Deliverables:

1. Code Data from EDA (Visual + statistics), data wrangling and model choice.
2. Report summarizing steps for each of the sections mentioned in step 1.
3. Presentation explaining each of the sections mentioned in step 1.