

Problem

Problem Statement

The Bank Indessa has not done well in the last 3 quarters. Their NPAs (Non Performing Assets) have reached all time high. It is starting to lose the confidence of its investors. As a result, its stock has fallen by 20% in the previous quarter alone.

After careful analysis, it was found that the majority of NPA was contributed by loan defaulters. With the messy data collected over all the years, this bank has decided to use machine learning to figure out a way to find these defaulters and devise a plan to reduce them.

This bank uses a pool of investors to sanction their loans. For example: If any customer has applied for a loan of \$20000, along with the bank, the investors perform a due diligence on the requested loan application. Keep this in mind while understanding data.

In this challenge, you will help this bank by predicting the probability that a member will default.

Data Information

There are files given: train, test and submission. Your submission file must adhere to the format specified in the given submission file. This data set comprises information captured in December 2016. Following is the description of variables given:

Variable	Description
member_id	unique ID assigned to each member

loan_amnt	loan amount (\$) applied by the member
funded_amnt	loan amount (\$) sanctioned by the bank
funded_amnt_inv	loan amount (\$) sanctioned by the investors
term	term of loan (in months)
batch_enrolled	batch numbers allotted to members
int_rate	interest rate (%) on loan
grade	grade assigned by the bank
sub_grade	grade assigned by the bank
emp_title	job / Employer title of member
emp_length	employment length, where 0 means less than one year and 10 means ten or more years
home_ownership	status of home ownership
annual_inc	annual income (\$) reported by the member

verification_status	status of income verified by the bank
pymnt_plan	indicates if any payment plan has started against loan
desc	loan description provided by member
purpose	purpose of loan
title	loan title provided by member
zip_code	first three digits of area zipcode of member
addr_state	living state of member
dti	ratio of member's total monthly debt repayment excluding mortgage divided by self reported monthly income
delinq_2yrs	number of 30+ days delinquency in past 2 years
inq_last_6mths	number of inquiries in last 6 months
mths_since_last_delinq	number of months since last delinq

mths_since_last_record	number of months since last public record
open_acc	number of open credit line in member's credit line
pub_rec	number of derogatory public records
revol_bal	total credit revolving balance
revol_util	amount of credit a member is using relative to revol_bal
total_acc	total number of credit lines available in members credit line
initial_list_status	unique listing status of the loan - W(Waiting), F(Forwarded)
total_rec_int	interest received till date
total_rec_late_fee	Late fee received till date
recoveries	post charge off gross recovery
collection_recovery_fee	post charge off collection fee

collections_12_mths_ex_med	number of collections in last 12 months excluding medical collections
mths_since_last_major_delinq	months since most recent 90 day or worse rating
application_type	indicates when the member is an individual or joint
verification_status_joint	indicates if the joint members income was verified by the bank
last_week_pay	indicates how long (in weeks) a member has paid EMI after batch enrolled
acc_now_delinq	number of accounts on which the member is delinquent
tot_coll_amt	total collection amount ever owed
tot_cur_bal	total current balance of all accounts
total_rev_hi_lim	total revolving credit limit
loan_status	status of loan amount, 1 = Defaulter, 0 = Non Defaulters

Evaluation Metric

Submissions will be evaluated based on AUC-ROC score and AUC-PR score