Loan Delinquency Prediction

Problem Statement

Loan default prediction is one of the most critical and crucial problem faced by financial institutions and organizations as it has a noteworthy effect on the profitability of these institutions. In recent years, there is a tremendous increase in the volume of non – performing loans which results in a jeopardizing effect on the growth of these institutions.

Therefore, to maintain a healthy portfolio, the banks put stringent monitoring and evaluation measures in place to ensure timely repayment of loans by borrowers. Despite these measures, a major proportion of loans become delinquent. Delinquency occurs when a borrower misses a payment against his/her loan.

Given the information like mortgage details, borrowers related details and payment details, our objective is to identify the delinquency status of loans for the next month given the delinquency status for the previous 12 months (in number of months)

Data Description

train.zip

train.zip contains train.csv. train.csv contains the training data with details on loan as described in the last section

Data Dictionary

Variable	**Description**
loan_id	Unique loan ID
source	Loan origination channel
financial_institution	Name of the bank
interest_rate	Loan interest rate
unpaid_principal_bal	Loan unpaid principal balance
loan_term	Loan term (in days)
origination_date	Loan origination date (YYYY-MM-DD)
first_payment_date	First instalment payment date
loan_to_value	Loan to value ratio
number_of_borrowers	Number of borrowers
debt_to_income_ratio	Debt-to-income ratio
borrower_credit_score	Borrower credit score
loan_purpose	Loan purpose
insurance_percent	Loan Amount percent covered by insurance
co-borrower_credit_score	Co-borrower credit score
insurance_type	0 - Premium paid by borrower, 1 - Premium paid by Lender
m1 to m12	Month-wise loan performance (deliquency in months)
m13	target, loan deliquency status (0 = non deliquent, 1 = deliquent)

test.zip

test.zip contains test.csv which has details of all loans for which the participants are to submit the delinquency status - 0/1 (not probability)

sample_submission.zip

sample_submission.zip contains the submission format for the predictions against the test set. A single csv needs to be submitted as a solution.

Evaluation Metric

Submissions are evaluated on F1-Score between the predicted class and the observed target.

Public and Private Split

Test data is further randomly divided into Public (40%) and Private (60%) data.

- Your initial responses will be checked and scored on the Public data.
- The final rankings would be based on your private score which will be published once the competition is over.

Hackathon Rules

1. Setting the final submission is mandatory. Without a final submission, the submission corresponding to best public score will be taken as final submission
2. Use of external datasets is not allowed
3. Use of loan_id variable as an input to the model is not allowed
4. You can only make 15 submissions per day
5. Code file is mandatory while setting final submission. For GUI based tools, please upload a zip file of snapshots of steps taken by you, else upload code file.
6. The code file uploaded should be pertaining to your final submission.
7. No submission will be accepted after the contest deadline