"Featured Prediction – Loan credit default risk"

Problem statement:

Can you predict how capable each applicant is of repaying a loan?

Bank "ABC" is challenged with loan defaulters. It is struggling with taking a decision on whether the client is capable to repay the loan or not.

Bank is using basic statistical models to understand the pattern in the data collected in the last few years.

Bank needs help in exploring the data to the fullest potential .Doing so will ensure that clients capable of repayment are not rejected and that loans will empower their clients to be successful.

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

Data Information:

Two files will be provided: (through git repository or Kaggle)

1) Credit_history_train.csv
2) Credit_history_test.csv

Github link - https://github.com/IBMDevConnect/RBS2018

(Sample data parameters)

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

Variable	Description
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

Variable	Description
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

Variable	Description
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_derog	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

Variable	Description
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:

- Based on AUC-ROC score
- Based on approach -> Feature engineering , Tools used

Files to upload:

1) Predicted value – csv file
2) Source file – file(s) describing the approach , feature engineering, tools and source code. Files to be compressed and uploaded to the git repository

Pre-requisite:

- Registration to IBM cloud (https://bluemix.net)
- Individual github repository from the participants to upload their source files

Skills:

- Statistical programming and basic ML concepts
- Awareness of various ML libraries , frameworks and tools
- Language awareness – python

Timeline: (3 weeks)

- ML hands on workshop and challenge announcement
    - One day
- Challenge
    - 2 weeks
- Evaluation
    - 1 week

Collaboration channel:

- Slack channel will be provided

Resources:

- https://developer.ibm.com/code/patterns/
- https://dataplatform.ibm.com/exchange/public/entry/view/fe3f5db21152f9e668009cb2f40cb096
- https://dataplatform.ibm.com/community?context=analytics&format=notebook
- IBM cloud resources – bluemix.net

Note: Resources specific to certain IBM cloud resources will be shared during the hands on workshop