MLZilla Case

Credit scoring and loan sanctioning are important processes in the lending industry. Credit scoring is a way of assessing the creditworthiness of an individual or business, based on their credit history and other financial information. A credit score is a numerical value that represents an individual's creditworthiness, with higher scores indicating less risk. Credit scoring is used by lenders to decide whether to approve a loan application and on what terms.

Loan sanctioning, also known as loan underwriting, is the process of evaluating a loan application and determining whether to approve or reject it. This process includes assessing the borrower's creditworthiness, as well as verifying their income, employment, and assets. The lender will also consider other factors such as the purpose of the loan and the amount of collateral provided. Once a loan application has been approved, the lender will set the terms of the loan, including the interest rate and repayment schedule.

Machine learning is increasingly being used in both credit scoring and loan sanctioning. By using large amounts of historical data, ML algorithms can learn to identify patterns that predict the likelihood of loan default and also help in evaluating the creditworthiness of the borrower. This analytical case study is an open-ended problem for you. Apart from the two problems mentioned, participants have a free hand to provide any other solution using the given dataset, the output of which can benefit the financial institution in question. Participants are free to use any data modeling/machine learning tools and techniques, no specific variable set has been mentioned as independent or dependent, it is up to your own approach, creativity, choice, idea, and analysis, as long as it can be justified with proper explanations. The description of the dataset is given below. The participants are free to make any logical assumptions.

What are we looking for?

* A well-defined problem statement on which the participants have based their solution on.
* The approaches taken for data cleaning and preprocessing.
* Deciding on the algorithms that will be used to solve the problem.
* Conversion of thoughts into a functioning code.
* Techniques used to validate the model(s).
* The business implication of the solution proposed.

Case solution submission guidelines:

* The case solution needs to be submitted in the form of a presentation.
* The presentation should contain the links to the code base, visualization tools or any other tools used to solve the problem.
* Each step of the case solution carries some weightage in terms of evaluation of the case. Therefore, participants will be marked even if the final solution is not reached.
* The marks will be awarded based on the creativity, the approach chosen, the reasoning behind it, models considered and the effectiveness of the presentation.
* The quality of insights arrived up on will also be an evaluating factor.
* Submission of the case should be done by the 21st of January 2023, Saturday, by 7 am.

## Data Description

Legend: LC=The company providing the loans

| Sl. No. | Variable Name | Description |
| --- | --- | --- |
| 1 | id | To uniquely identify every loan in the dataset. |
| 2 | member\_id | To identify the borrower to who has applied for the loan. |
| 3 | loan\_amnt | The listed amount of the loan applied for by the borrower. |
| 4 | funded\_amnt | The amount that was sanctioned by the LC. |
| 5 | term | The number of payments on the loan. Values are in months and can be either 36 or 60. |
| 6 | int\_rate | Interest Rate on the loan |
| 7 | installment | The monthly payment owed by the borrower if the loan originates. |
| 8 | grade | LC assigned loan grade which depends on the borrower’s credit score. |
| 9 | sub\_grade | LC assigned loan subgrade |
| 10 | emp\_title | The job title supplied by the Borrower when applying for the loan.\* |
| 11 | emp\_length | Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years. |
| 12 | home\_ownership | The home ownership status provided by the borrower during registration or obtained from the credit report. Our values are: RENT, OWN, MORTGAGE, OTHER |
| 13 | annual\_inc | The self-reported annual income provided by the borrower during registration. |
| 14 | verification\_status | Indicates if income was verified by LC, not verified, or if the income source was verified |
| 15 | issue\_d | The month which the loan was funded |
| 16 | loan\_status | Current status of the loan |
| 17 | purpose | A category provided in the form of a code to indicate the purpose for the loan. |
| 18 | title | Explaining the ‘purpose’ of the loan. |
| 19 | dti | The debt to income ratio is the ratio of how much the borrower owes every month to the borrower’s income every month. |
| 20 | delinq\_2yrs | The number of delinquencies(late installment payment) by the borrower in the past 2 years. |
| 21 | earliest\_cr\_line | The month-year the borrower's earliest reported credit line was opened |
| 22 | inq\_last\_6mths | Inquiries for loans made by the borrower over the past 6 months. |
| 23 | mths\_since\_last\_delinq | Months that have passed since the borrower last missed the timely payment of installment. |
| 24 | open\_acc | The number of open credit lines in the borrower’s credit file. |
| 25 | pub\_rec | Number of derogatory public records |
| 26 | revol\_bal | Total credit revolving balance |
| 27 | revol\_util | Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit. |
| 28 | total\_acc | The total number of credit lines currently in the borrower's credit file |
| 29 | initial\_list\_status | The initial listing status of the loan. Possible values are – W(whole), F(fractional) |
| 30 | out\_prncp | Remaining outstanding principal for total amount funded |
| 31 | total\_pymnt | Payments received to date for the total amount funded. |
| 32 | total\_rec\_prncp | Principal received till date. |
| 33 | total\_rec\_int | Interest received till date. |
| 34 | total\_rec\_late\_fee | Late fees received to date. |
| 35 | recoveries | Total recovery procedures initiated against the borrower. |
| 36 | collection\_recovery\_fee | The fees collected during the recovery procedures. |
| 37 | last\_pymnt\_d | The last month when payment was received. |
| 38 | last\_pymnt\_amnt | The last payment amount received. |
| 39 | next\_pymnt\_d | Next scheduled payment date. |
| 40 | last\_credit\_pull\_d | The most recent month LC pulled credit for this loan |
| 41 | collections\_12\_mths\_ex\_med | Number of collections in 12 months excluding medical collections |
| 42 | mths\_since\_last\_major\_derog | Months since most recent 90-day delinquency or worse rating |
| 43 | application\_type | Indicates whether the loan is an individual application or a joint application with two co-borrowers |
| 44 | annual\_inc\_joint | The combined self-reported annual income provided by the co-borrowers during registration |
| 45 | dti\_joint | A ratio calculated using the co-borrowers' total monthly payments on the total debt obligations, excluding mortgages and the requested LC loan, divided by the co-borrowers' combined self-reported monthly income |
| 46 | acc\_now\_delinq | The number of accounts on which the borrower is now delinquent |
| 47 | tot\_coll\_amt | Total collection amounts ever owed by the borrower |
| 48 | tot\_cur\_bal | Total current balance of all accounts owned by the borrower |
| 49 | total\_rev\_hi\_lim | Total high credit/credit limit |