**Credit Risk Assessment in Fintech Industry**

**Problem Statement:**

Welcome to Inceptez hackathon – **Credit Risk in Fintech Industry.** You are required to build and train a model that identifies Fully Paid and Charged-off loans from the loan dataset. This dataset is included in loan data, and provides a challenging classifier that will test what you have learnt in this course.

**Task:**

Your task is to build this model based on the details in this document and submit it. Please read the details carefully before attempting this hackathon.

You will need to decide the following:

1. Use the specific source or dataset for assess credit risk shared with you
2. What is your intended data split ratio for training, validation, and test sets for the loan dataset? How do you plan to ensure randomness in this split?
3. Do you plan to explore the importance of these components further?
4. Do you anticipate class imbalance in the 'loan\_status' feature, where

*Fully paid: Applicant has fully paid the loan (the principal and the interest rate)*

*Charged-off: Applicant has not paid the installments in due time for a long period of time, i.e. Client has defaulted on the loan*

If so, how will you address this imbalance?

1. Will you normalize the features? If yes, what normalization techniques do you have in mind?
2. Do you intend to perform data preprocessing tasks such as outlier detection, missing value handling, or feature selection before training your model.

**Dataset:**

Although the dataset is provided with limited columns you have some decisions to make around it. First, it's important to familiarize yourself with the dataset:

You will note that :

There are 80,000 records and 28 features and data description is given below

|  |  |
| --- | --- |
| **Column** | **Description** |
| earliest\_cr\_line | The month the borrower's earliest reported credit line was opened |
| emp\_title | The job title supplied by the Borrower when applying for the loan. |
| fico\_range\_high | The upper boundary range the borrower’s FICO at loan origination belongs to. |
| fico\_range\_low | The lower boundary range the borrower’s FICO at loan origination belongs to. |
| grade | LC assigned loan grade |
| application\_type | Indicates whether the loan is an individual application or a joint application with two co-borrowers |
| initial\_list\_status | The initial listing status of the loan. Possible values are – W, F |
| num\_actv\_bc\_tl | Number of currently active bankcard accounts. |
| mort\_acc | Number of mortgage accounts. |
| tot\_cur\_bal | Total current balance of all accounts |
| open\_acc | The number of open credit lines in the borrower's credit file. |
| pub\_rec | Number of derogatory public records |
| pub\_rec\_bankruptcies | Number of public record bankruptcies. |
| purpose | A category provided by the borrower for the loan request. |
| revol\_bal | Total credit revolving balance |
| title | The loan title provided by the borrower |
| total\_acc | The total number of credit lines currently in the borrower's credit file |
| verification\_status | Indicates if income was verified by LC, not verified, or if the income source was verified |
| addr\_state | The state provided by the borrower in the loan application |
| annual\_inc | The self-reported annual income provided by the borrower during registration. |
| 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. |
| home\_ownership | The home ownership status provided by the borrower during registration. Our values are: RENT, OWN, MORTGAGE, OTHER. |
| int\_rate | Interest Rate on the loan |
| loan\_amnt | The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value. |
| loan\_status | Current status of the loan |
| sub\_grade | LC assigned loan subgrade |
| term | The number of payments on the loan. Values are in months and can be either 36 or 60. |
| revol\_util | Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit. |

**Submission:**

Please submit the Jupyter Notebook. This should clearly show:

1. The code you have written

2. The output of the code

3. A comprehensive description of each code block, together with the decisions you've made and the rationale for those decisions.

4. A comprehensive description of what you also tried that did not work, and what lessons you have learnt from this hackathon.

5. Finally, prepare a PowerPoint presentation show casing your project

**Note:**

﻿• Use Python programming language

• Use a Google Colab, or a standard laptop/desktop to build the model

• This is a challenging dataset to predict accurately on, so iterate your approach over time making note of what works and what doesn't. In the long term, this is as useful to you as getting a high model accuracy.

• Avoid Plagiarism as the objective of this exercise is to give you a real-world project to build and we hope you will use this opportunity wisely to your benefit.