

Modelling Test

Congratulations on making it to the Monsoon modelling test. Making it this far means you have already worked through a rigorous coding test and tough interview. This round will test your ability to build models from raw data. We will evaluate your submission on model performance and the process followed to build the model.

Problem Statement

This dataset has been provided by your client, a lending company. They have provided application details of all previous loans that they have disbursed. This data contains information collected at the time of application and a marker to identify a loan as a default case or a loan that was paid on time.

You are asked to prepare a model that can predict the loan outcome at the time of application.

Data Provided

The data folder contains data for both the train and test sets. Both train and test folders will contain data accounts data and enquiry data. These datasets are provided as JSON lists. Both files also include a flag csv. This dataset is provided as a csv file.

Brief descriptions of the data are below and details of the fields are later in this document.

1. Train Data: This contains details of the loan provided by your client.
2. Accounts Data: Details of loans taken by the borrower *before* they applied for a loan from your client. These details were collected when the borrower applied for a loan from your client.
3. Enquiry Data: This contains details of previous loan applications made by the applicant. These applications may or may not be successful.

Instructions

Please write your final prediction to the final_submission folder in the test provided. Please save the file in the same format as the sample_submission.csv. The filename should be final_submission_'your_first_name_your_last_name'.csv



The evaluation metric will be the roc_auc_score.

Also include all your code and notebooks in the submission link. We want to see how you conducted EDA, feature creation, feature selection and modelling attempts along with the final score.

Flag Data Details

Field Name	Explanation
uid	A unique ID used to identify the applicant
NAME_CONTRACT_TYPE	Type of loan that has been given to the borrower
TARGET	Boolean field that represents the outcome of the loan. 1 = bad loan 0 = good loan

Accounts Data Details

Field Name	Explanation
uid	A unique ID used to identify the applicant
credit_type	The type of the loan.
loan_amount	The principal amount of the loan
amount_overdue	The amount of the loan that is currently passed due date.
open_date	Date on which the loan was sanctioned
closed_date	Date on which the loan was closed. If the loan is running, then this is missing.

Field Name	Explanation
payment_hist_string	<p>A month by month repayment history for that loan.</p> <p>Three digits represent the payment history for a single month. The numeric value represents the number of days overdue for that month.</p> <p>The most recent payment is on the right of the string.</p> <p>For example, if the string is: '000026' then the borrower was 26 days past the due date when making the payment for the most recent month, but was on-time (0 days past due) for the month before that.</p>

Enquiry Data Details

Field Name	Explanation
uid	A unique ID used to identify the applicant
enquiry_type	The type of the loan that was applied for.
enquiry_amt	The request loan amount.
enquiry_date	The date on which the application was made.