Credit Risk Analysis (P2P lending platform)

Target Variables Creation Report

Platform: Bondora

By: Nour M. Ibrahim

As a Target Variable for the Regression Model, There are 3 variables have been requested to formulate:

1- Equated Monthly Installments (EMI)
2- Eligible Loan Amount (ELA)
3- Preferred Return On Investment (PROI)

This is report is aimed to illustrate the approach, methodology, and procedure of creating these target variables from the preexistent independent variables in the dataset.

I. Loan Tenure

Loan Tenure, or the number of months starting from the loan issue date until its maturity date, is a key factor in calculating all 3 target variables,

- Components of Loan Tenure:
  - LoanDate
  - MaturityDate_Original
- **Calculation Procedure**:

  LoanTenure = $\left(MaturityDat\_Original_{year} - LoanDate_{year}\right) x\ 12 - (MaturityDate\_Original_{month} - LoanDate_{month})$

II. Equated Monthly Installments (EMI)

- EMI refers to the fixed amount of money the borrower pays to a lender as part of the repayment towards an outstanding loan within a special period.
- Components of EMI is
  - Tenure
  - Principle repayment
  - Interest
- There's more than one method for calculating EMI, but we'll be using the Flat rate method, which follows this formula,

  EMI = P x r x $\frac{(1+r)^n}{((1+r)^n-1)}$

  In our dataset:
  P: "Amount" in a month
  R: "Interest"
  N: "LoanTenure"
- **Calculation Procedure**:

  For each row in the dataset:
  1- Calculate result_1 = P * r * $(1+r)^n$
  2- Calculate result_2 = $(1+r)^n - 1$
  3- Calculate EMI = result_1 / result_2

III. Eligible Loan Amount (ELA)

- ELA refers to the amount of loan that a borrower is eligible for after reviewing his data.
- ELA should reflect the amount a lender is willing to grant the borrower.
- Components of ELA:
  - A: "AppliedAmount"
  - R: "Interest"
  - N: "LoanTenure"
  - I: "IncomeTotal"
  - L: "LiabilitiesTotal"
- **Calculation Procedure:**

  For each row in the dataset:

  1- Calculate: Total Payment Due = (A + (A*r) * n

  2- Calculate: **Max allowable amount = (I – L) * 30%**

  3- If (Total Payment Due <= Max allowable amount)

  Then ELA = AppliedAmount

  Else ELA = Max allowable amount

IV. Preferred Return on Investment (PROI)

- Return on Investment or ROI is a popular profitability metric used to evaluate how well an investment has performed, It tries to directly measure the amount of return on a particular investment.
- ROI is calculated using many formulas depending on the case in hand, one of them is

  ROI = $\frac{\boldsymbol{Interest\ Amount}}{\boldsymbol{Loan\ Amount}}\ \boldsymbol{x\ 100}$
- In order to account for the risk of investing, we used a feature importance technique, followed by the weight of evidence technique to select the highly effective independent variables, and which interval inside each variable has much wait in classifying a default or non-default loan, The method is as follows:

  <u>Feature Importance Selection:</u>

  1- Using RandomForestRegressor Model from scikit-learn library in Python, we Identified the these Independent variables as the most important while predicting ROI,
     a- Interest
     b- AppliedAmount
     c- LoanTenure
     d- IncomeTotal
     e- LiabilitiesTotal
     f- DebtToIncome

  2- Since Interest is used in the calculation of ROI, then we'll exclude it, and Since DebtToIncome is covering LiabilitiesTotal information especially when I keep IncomeTotal, Then I'll exclude LiabilitiesTotal as well.

  <u>Weight of Evidence and Information value:</u>

  Using the remaining 4 independent variables, I'll start a procedure for determining the effectiveness of specific intervals inside each variable on the LoanStatus, whether the borrower defaulted or not,

  For each one of the 4 variables,

  1- Divide the range of the variables in 50 equal ranges starting of min value and ending at max.

  2- Calculate WoE for each interval using this equation

$$\boldsymbol{WoE_i = \ln\left(\frac{\%(LoanStatus = 1)_i}{\%(LoanStatus = 0)_i}\right)}$$

And Then Calculate the Information Value on the independent variable using this equation,

IV = $\sum_{i=1}^{50}(P(LoanStatus = 1) - P(LoanStatus = 0))\ x\ WoE_i$

3- Redefine the intervals of each variable depending on the results of WoE.
4- Choose which intervals reduces the risk of default, which would increase it, and which ones are relatively neutral, I came up with these results,

Independent Variable	Interval	Effect on risk of default
LoanTenure	≤ 19 months	decrease risk
LoanTenure	> 25 months	Increase risk
AppliedAmount	850 ≤ amnt ≤ 1175	decrease risk
AppliedAmount	amnt ≥ 2000	increase risk
IncomeTotal	amnt ≤ 1000	decrease risk
DebtToIncome	rate = 0	decrease risk
DebtToIncome	rate ≠ 0	increase risk

All other intervals have neutral effect on a Loan being defaulted or not.

Preferred ROI (PROI) Calculation Procedure:

After defining the intervals that has effect on risk of default of the loan, I used the following procedure to calculate the risk of value,

1- Calculate ROI = $\frac{\text{InterestAmount}}{\text{Amount}}\ x\ 100$ for all rows on the dataset

For each row in the dataset,

2- Set **PROI = median(ROI)** of the all values in the dataset **(anchor value)**
3- Check AppliedAmount,
   If 850 ≤ AppliedAmount ≤ 1175:
      PROI = PROI – 5 (decrease ROI with 5%)
   If AppliedAmount ≥ 2000:
      PROI = PROI + 5 (increase ROI with 5%)
4- Check for LoanTenure,
   If LoanTenure ≤ 19:
      PROI = PROI – 5 (decrease ROI with 5%)
   If LoanTenure > 25:
      PROI = PROI + 5 (increase ROI with 5%)
5- Check IncomeTotal:
   If IncomeTotal ≤ 1000:
      PROI = PROI – 5 (decrease ROI with 5%)
6- Check for DebtToIncome:
   If DebtToIncome == 0:
      PROI = PROI – 5 (decrease ROI with 5%)
   Else:
      PROI = PROI + (Increase ROI with 5%)
7- Return final PROI.

This concludes the report.

Best Regards…