Business Problem

ABC Customer Bank is a mid-sized private bank which deals in all kinds of loans. They have presence across all major cities in India and focus on lending products. They have a digital arm which sources customers from the internet.

Digital arms of banks today face challenges with lead conversion, they source leads through mediums like search, display, email campaigns and via affiliate partners. Here ABC Customer Bank faces same challenge of low conversion ratio. They have given a problem to identify the customers segments having higher conversion ratio for a specific loan product so that they can specifically target these customers, here they have provided a partial data set for salaried customers only from the last 3 months. They also capture basic details about customers like gender, DOB, existing EMI, employer Name, Loan Amount Required, Monthly Income, City, Interaction data and many others. Let's look at the process at ABC Customer Bank.

In above process, customer applications can drop majorly at two stages, at login and approval/ rejection by bank. Here we need to identify the segment of customers having higher disbursal rate in next 30 days.

DATA AVAILABLE:
We have train and test data set, train data set has both input and output variable(s). Need to predict probability of disbursal for test data set.

DATA DESCRIPTION:

ID - Unique ID (can not be used for predictions)
Gender- Sex
City - Current City
Monthly_Income - Monthly Income in rupees
DOB - Date of Birth

Lead_Creation_Date - Lead Created on date
Loan_Amount_Applied - Loan Amount Requested (INR)
Loan_Tenure_Applied - Loan Tenure Requested (in years)
Existing_EMI - EMI of Existing Loans (INR)
Employer_Name - Employer Name
Salary_Account- Salary account with Bank
Mobile_Verified - Mobile Verified (Y/N)
Var5- Continuous classified variable
Var1- Categorical variable with multiple levels
Loan_Amount_Submitted- Loan Amount Revised and Selected after seeing Eligibility
Loan_Tenure_Submitted- Loan Tenure Revised and Selected after seeing Eligibility (Years)
Interest_Rate- Interest Rate of Submitted Loan Amount
Processing_Fee- Processing Fee of Submitted Loan Amount (INR)
EMI_Loan_Submitted- EMI of Submitted Loan Amount (INR)
Filled_Form- Filled Application form post quote
Device_Type- Device from which application was made (Browser/ Mobile)
Var2- Categorical Variable with multiple Levels
Source- Categorical Variable with multiple Levels
Var4- Categorical Variable with multiple Levels

OUTCOMES:

LoggedIn- Application Logged (Variable for understanding the problem – cannot be used in prediction)
Disbursed- Loan Disbursed (Target Variable)

Evaluation Cirteria:

Evaluation metrics: ROC_AUC, CONCORDANCE, SENSITIVITY, SPECICITY etc