Case Study
on
Bank Loan Defaulter Insights

Team:

Vasantha Raj Tejas

Kumar

Final Presentation

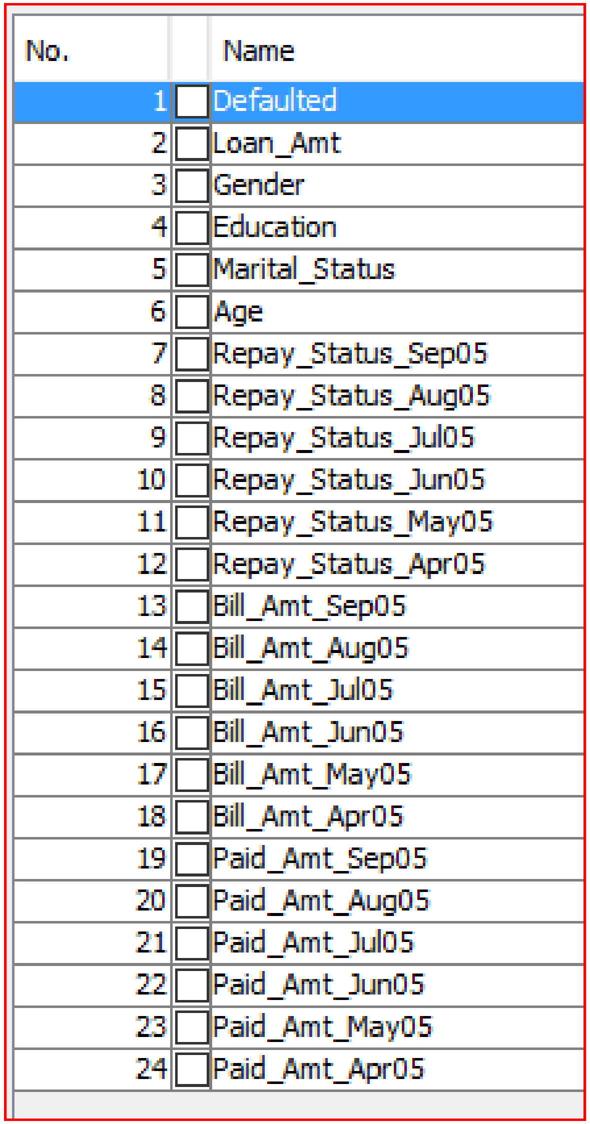
About Case Study

- Banks need to identify good customers from bad customers who default on their loans from time to time in order to optimize their loan disbursal and recovery.
- Herein, an appropriate customer selection from their available demographics and other related information from a prospect helps banks reduce to defaulters either in front or target with some promotional offers that of responsorial in nature and avoid defaulting.
- In order to understand who will be (or characteristics that explain) defaulters, a set of classification and regression tree models were built to get analytical insights.
- Thus, study proceeded with a business objective, “to identify the drivers of loan defaulters based on a set of independent (features) variables”.

Data Employed for the Case Study

We had total of 24 variables in the obtained file, such as:

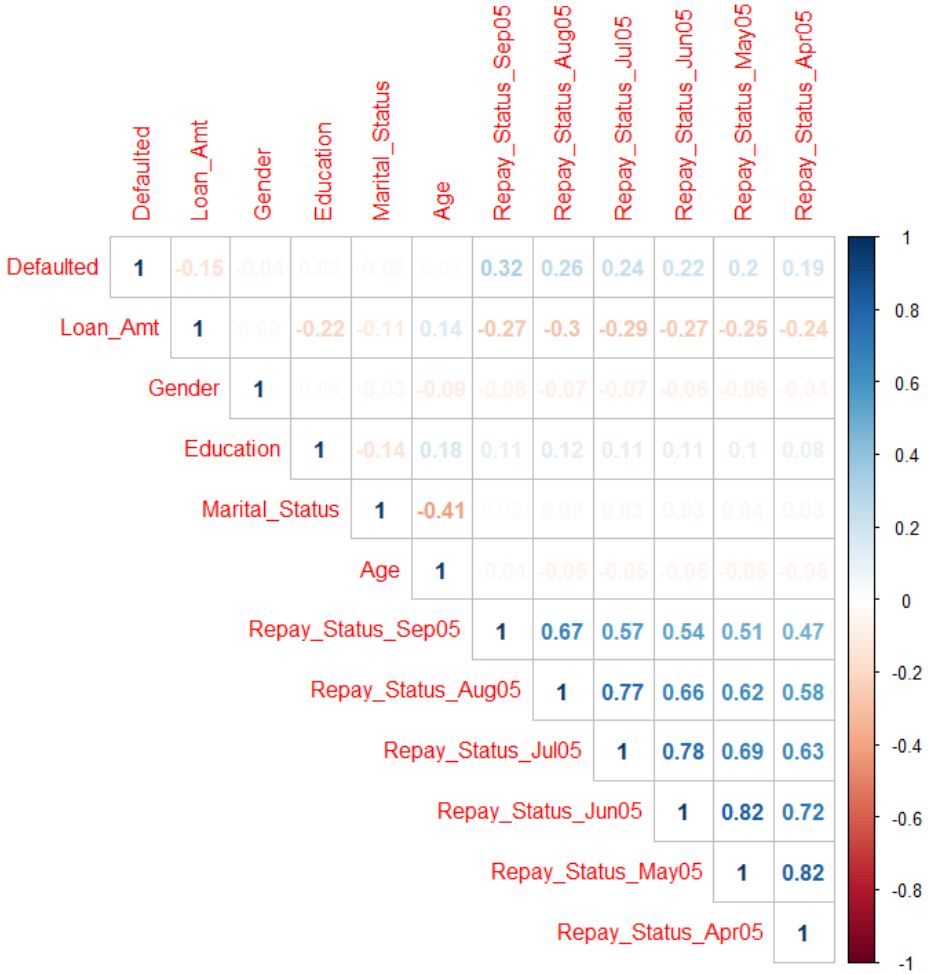
- Defaulted – Target - if customer defaulted payment it takes value of 1, otherwise 0.
- Loan Details:
  - Loan_Amt - Amount of the given credit (in dollars).
  - Repay_Status_xxxxx (6 variables) – Whether customer repaid for the said month.
  - Bill_Amt_xxxxx (6 variables) – Customers installment amount to be paid for the said month.
  - Paid_Amt_xxxxx (6 variables – How much amount customer paid in the last month.
- Demographics:
  - Gender Details,
  - Customer Education Details
  - Customer Age
  - Marital Status



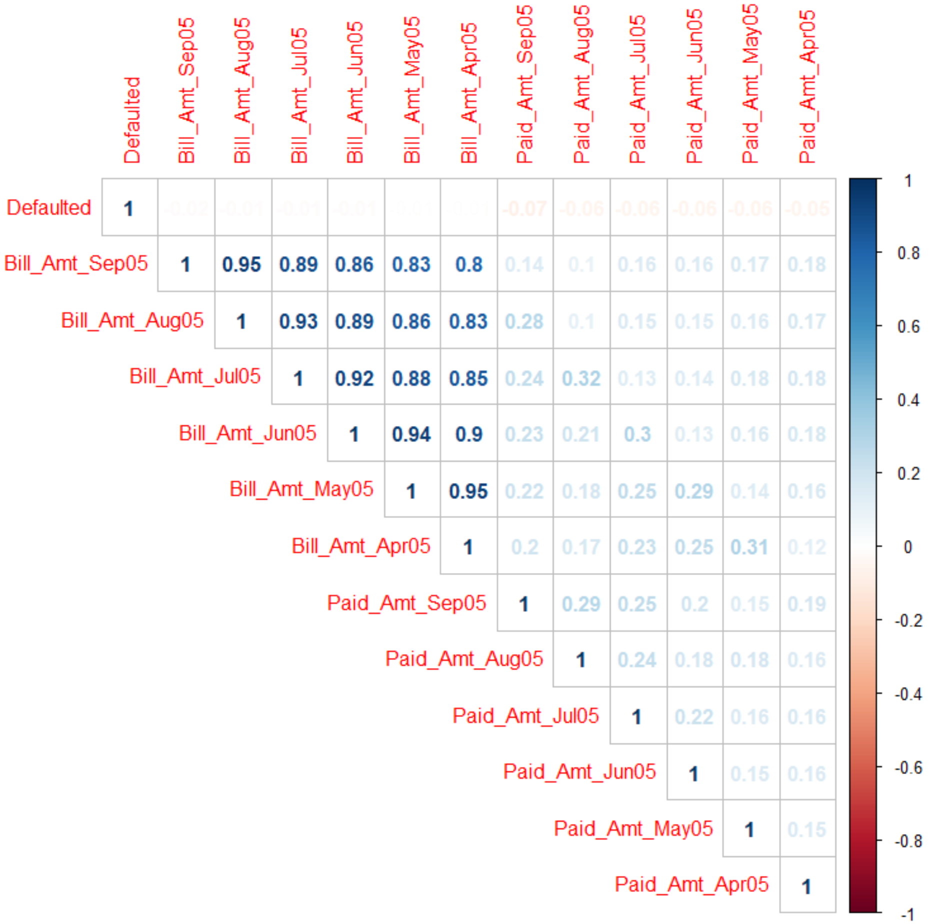
No.		Name
1	□	Defaulted
2	□	Loan_Amt
3	□	Gender
4	□	Education
5	□	Marital_Status
6	□	Age
7	□	Repay_Status_Sep05
8	□	Repay_Status_Aug05
9	□	Repay_Status_Jul05
10	□	Repay_Status_Jun05
11	□	Repay_Status_May05
12	□	Repay_Status_Apr05
13	□	Bill_Amt_Sep05
14	□	Bill_Amt_Aug05
15	□	Bill_Amt_Jul05
16	□	Bill_Amt_Jun05
17	□	Bill_Amt_May05
18	□	Bill_Amt_Apr05
19	□	Paid_Amt_Sep05
20	□	Paid_Amt_Aug05
21	□	Paid_Amt_Jul05
22	□	Paid_Amt_Jun05
23	□	Paid_Amt_May05
24	□	Paid_Amt_Apr05

Exploratory Data Analysis

Correlation Plot of the Data



	Defaulted	Loan_Amt	Gender	Education	Marital_Status	Age	Repay_Status_Sep05	Repay_Status_Aug05	Repay_Status_Jul05	Repay_Status_Jun05	Repay_Status_May05	Repay_Status_Apr05
Defaulted	1	-0.15	-0.04	0.03	-0.02	0.01	0.32	0.26	0.24	0.22	0.2	0.19
Loan_Amt		1	0.02	-0.22	-0.11	0.14	-0.27	-0.3	-0.29	-0.27	-0.25	-0.24
Gender			1	0.01	-0.03	-0.09	-0.06	-0.07	-0.07	-0.06	-0.06	-0.04
Education				1	-0.14	0.18	0.11	0.12	0.11	0.11	0.1	0.08
Marital_Status					1	-0.41	0.02	0.02	0.03	0.03	0.04	0.03
Age						1	-0.04	-0.05	-0.05	-0.05	-0.05	-0.05
Repay_Status_Sep05							1	0.67	0.57	0.54	0.51	0.47
Repay_Status_Aug05								1	0.77	0.66	0.62	0.58
Repay_Status_Jul05									1	0.78	0.69	0.63
Repay_Status_Jun05										1	0.82	0.72
Repay_Status_May05											1	0.82
Repay_Status_Apr05												1



	Defaulted	Bill_Amt_Sep05	Bill_Amt_Aug05	Bill_Amt_Jul05	Bill_Amt_Jun05	Bill_Amt_May05	Bill_Amt_Apr05	Paid_Amt_Sep05	Paid_Amt_Aug05	Paid_Amt_Jul05	Paid_Amt_Jun05	Paid_Amt_May05	Paid_Amt_Apr05
Defaulted	1	-0.02	-0.01	-0.01	-0.01	-0.01	-0.01	-0.07	-0.06	-0.06	-0.06	-0.06	-0.05
Bill_Amt_Sep05		1	0.95	0.89	0.86	0.83	0.8	0.14	0.1	0.16	0.16	0.17	0.18
Bill_Amt_Aug05			1	0.93	0.89	0.86	0.83	0.28	0.1	0.15	0.15	0.16	0.17
Bill_Amt_Jul05				1	0.92	0.88	0.85	0.24	0.32	0.13	0.14	0.18	0.18
Bill_Amt_Jun05					1	0.94	0.9	0.23	0.21	0.3	0.13	0.16	0.18
Bill_Amt_May05						1	0.95	0.22	0.18	0.25	0.29	0.14	0.16
Bill_Amt_Apr05							1	0.2	0.17	0.23	0.25	0.31	0.12
Paid_Amt_Sep05								1	0.29	0.25	0.2	0.15	0.19
Paid_Amt_Aug05									1	0.24	0.18	0.18	0.16
Paid_Amt_Jul05										1	0.22	0.16	0.16
Paid_Amt_Jun05											1	0.15	0.16
Paid_Amt_May05												1	0.15
Paid_Amt_Apr05													1

Data Preparation & Feature Selection

- Required variables namely., Defaulted, Gender, Education, and Marital Status were converted to nominal.
- Applied the train & test at split ratio of 70:30.
- Employed "GLM Net' evaluator for feature selection process and selected below variables for modeling analysis.

Variable Rank	Variable Name
1	Loan_Amt - Customer Loan Amount
2	Age – Customer Age at the time of loan disbursal
3	Repay_Status_Sep05 – Last month repayment status
4	Education – Customer Educational Details
5	Bill_Amt_Sep05 – Installment amount to be paid.

Machine Learning Application

- Applied following Classification and Regression Trees Models:
  - As client was interested in also understanding power of each independent (feature), first, we attempted to classify using Binary Logistic Regression (BLR). However, it didn't yield good explanatory power, hence moved to decision trees.
  - Total four varieties of decision trees have been employed namely., C5.0, Boosting (ADA Boost), Bagging (Tree Bag) and RF (Random Forests) for obtaining best classifier for given data which will be used for scoring of future data sets.
  - All model accuracy and explanatory power have been reported duly in coming slides.

Interpretation of ML Application

- Employed 10-fold cross validation for all employed models. It is observed that C5.0 classifier has provided best accuracy at 71.1% against all others.
- Also, provided variable importance of all the features (independents) from the C5.0 model results.

Models	Accuracy (AUC)
Binary Logistic Regression (BLR)	0.603
C5.0	0.711
Ada Boost	0.677
Bagging Trees	0.687
Random Forests	0.697

Variable Name	Importance
Repay_Status_Sep05	1
Gender	2
Education	3
Loan Amount	4
Marital Status	5
Paid_Amt_Aug05	6

Business Insights

- Case Study – provided few analytical insights for banks which can be very helpful to identifying future defaulters and also while disbursing new loans.
- It is observed that customers having graduation or below as education more inclined to defaulters list.
- Further, all those customers who take large loan amounts (>90,000) and who age is below 24 years are prone to defaulting.
- Thus, insights helps banks to create efforts towards the campaign that target higher educational, and age customers to reduce the defaulting in the future.
- Also, while customers education back ground is graduation or less, banks can either stop providing loans to them or for only who are of greater than age of 25.