

Data Science PROJECT

Client: Bank GoodCredit | Category: Banking - Risk

Project Ref: PM-PR-0015

Business Case:

Bank GoodCredit wants to predict cred score for current credit card customers. The cred score will denote a customer's credit worthiness and help the bank in reducing credit default risk.

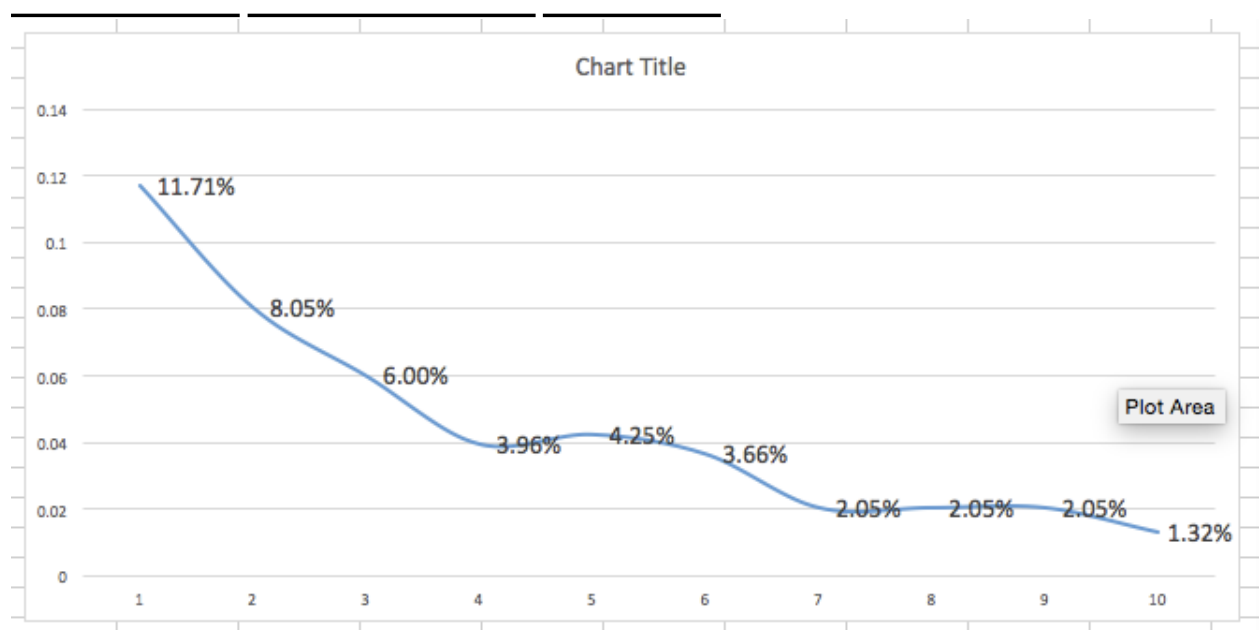
Target variable → Bad_label

0 – Customer has Good credit history

1 – Customer has Bad credit history (falls into 30 DPD + bucket)

Our benchmark model has gini as 37.9.

Below is our rank ordering



Decile Rank	66 features (Ensemble)
10	0.11713031
9	0.08052709
8	0.06002928
7	0.03958944
6	0.04245974
5	0.03660322
4	0.02052786
3	0.0204978
2	0.0204978
1	0.01317716

Data Explanation

Customer Account Data (Table : Cust_Account):

This table contains customer's historical accounts data and payments history

Customer Enquiry Data (Table : Cust_Enquiry):

This table contains customer's historical enquiry data such as enquiry amount and enquiry purpose.

Demographics Data (Table : Cust_Demographics):

Current customer applications with demographic data

Note that demographics features are renamed as features and obscured in accordance with privacy policies.

PROJECT GOAL:

Build a model with the data provided

1. Data exploration insights – what did you find and what decision did you take?
2. Feature matrix - List of features selected with gain
3. Model evaluation - Gini and rank ordering

Data Set Fields:

Host: 18.136.157.135

Port: 3306

Username : dm_team1

Password: DM!\$Team&279@20!

Database : project_banking

Customers Account Table

dt_opened
customer_no
upload_dt
acct_type
owner_indic
opened_dt
last_paymt_dt
closed_dt
reporting_dt
high_credit_amt
cur_balance_amt
amt_past_due
paymenthistory1
paymenthistory2
paymt_str_dt
paymt_end_dt
creditlimit
cashlimit
rateofinterest
paymentfrequency
actualpaymentamount

Enquiry Table

dt_opened
customer_no
upload_dt
enquiry_dt
enq_purpose
enq_amt

Customer Demographics Table

dt_opened

customer no

entry_time

feature_1 to feature_79

Bad_label

Some feature suggestions

Feature	Gain	Remarks
payment_history_avg_dpd_0_29_bucket	0.045653821	mean count of accounts that is in 0-29 dpd bucket throughout the payment history
total_diff_lastpaymt_opened_dt	0.04083407	The total duration between last payment date and account opened date of all accounts
min_months_last_30_plus	0.038216344	The smallest number of months passed before first 30+ dpd appeared for each account.
utilisation_trend	0.037523553	$\frac{[\text{total cur_bal_amt} / \text{total credit limit}]}{[\text{mean cur_bal_amt} / (\text{mean credit limit} + \text{mean_cashlimit})]}$
count_enquiry_recency_365	0.036216624	# of enquiry made in past 365 days
Ratio_currbalance_creditlimit	0.035640536	$[\text{total cur_bal_amt} / \text{total credit limit}]$
mean_diff_lastpaymt_opened_dt	0.032185579	The average duration between last payment date and account opened date of all accounts
mean_diff_open_enquiry_dt	0.031057926	average difference between enquiry dt_opened date and enquiry date
payment_history_mean_length	0.029258469	average length of payment_history variable
max_freq_enquiry	0.024501434	most frequent enquiry purpose
count_enquiry_recency_90	0.023796745	# of enquiry made in past 90 days
perc_unsecured_others	0.023428297	ratio of secured loan type enquiry purpose to total enquiry purpose made