

# 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.

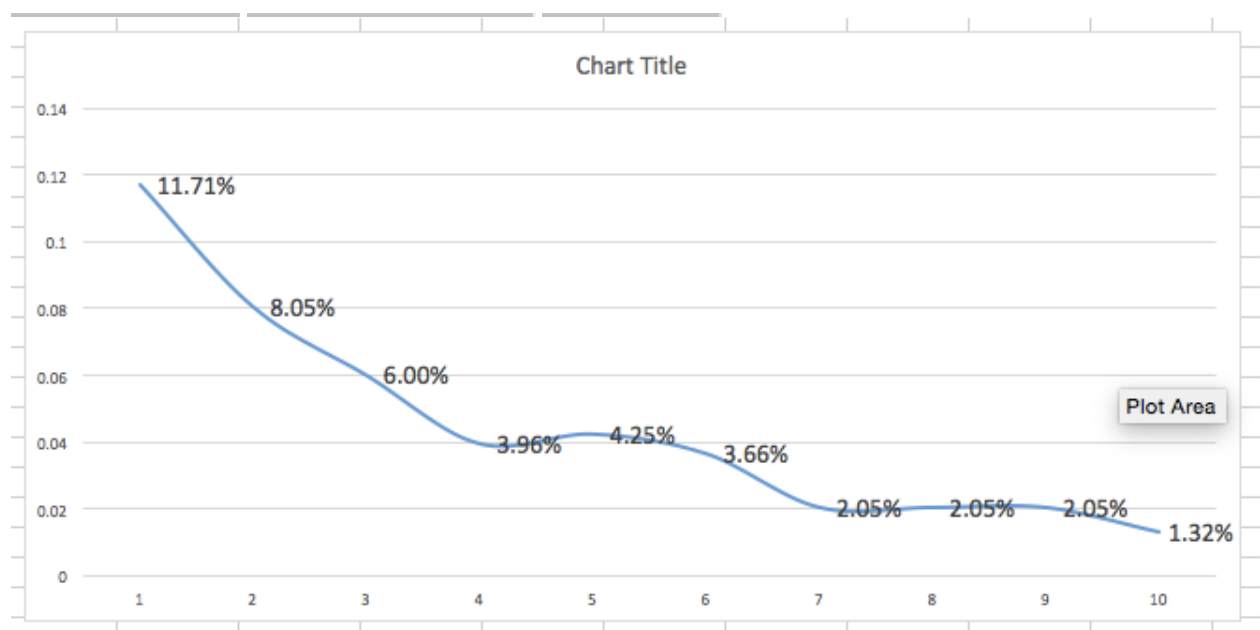
Target variable → Bad\_label

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: 18.136.56.185

Port: 3306

Username : dm\_team1

Password: dm\_team1123#

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	$[\text{total cur\_bal\_amt} / \text{total credit limit}] / [\text{mean cur\_bal\_amt} / (\text{mean credit limit} + \text{mean\_cashlimit})]$
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