

DS_Proj

Kai Yip Lei (Derek)

mars 19, 2017

1. Setup, Loading libraries and Data

I started the task by loading the libraries and data.

```
library(dplyr)
library(ggplot2)
library(RCurl)
library(rmarkdown)
library(knitr)
library(stats4)
library(stats)
library(pander)
library(NbClust)
library(randomForest)
library(ROCR)
library(glmnet)
library(e1071)
library(doParallel)
library(data.table)
library(tidyr)
library(gridExtra)

rm(list=ls())
dr<-data.table(read.csv("default of credit card clients.csv",header=TRUE,sep="," ,skip=1)%>%rename(y=default.payment.next.month))
set.seed(0706)

opts_chunk$set(cache=TRUE)
options(digit=4)
```

The response variable *Default Payment Next Month* was renamed to *y* for convenience and the seed was set to **0706** here to ensure reproducibility.

2. Exploratory data analysis & Feature Engineering

a. Data description

The variables in the dataset are defined as follows:

* Item 1 * Item 2 + Item 2a + Item 2b