Assignment3\_1

Sangamesh

18 September 2018

# Business Understanding

I’m a credit modeler in a bank and I’ve to develope a model that predicts the default probability of a loan.

# Data Mining Goals

DETERMINE BUSINESS OBJECTIVES

bankDetails <- read.csv("BankBayesLoan.txt", stringsAsFactors = FALSE,header=TRUE, sep="\t")  
trainData <- bankDetails[1:600,1:9]  
print(summary(trainData))

## age ed employ address   
## Min. :20.00 Min. :1.000 Min. : 0.00 Min. : 0.000   
## 1st Qu.:28.00 1st Qu.:1.000 1st Qu.: 3.00 1st Qu.: 3.000   
## Median :34.00 Median :1.000 Median : 7.00 Median : 7.000   
## Mean :34.66 Mean :1.745 Mean : 8.28 Mean : 8.203   
## 3rd Qu.:40.00 3rd Qu.:2.000 3rd Qu.:13.00 3rd Qu.:12.000   
## Max. :56.00 Max. :5.000 Max. :31.00 Max. :34.000   
## income debtinc creddebt othdebt   
## Min. : 14.00 Min. : 0.40 Min. : 0.0117 Min. : 0.04558   
## 1st Qu.: 24.00 1st Qu.: 5.00 1st Qu.: 0.3700 1st Qu.: 1.01287   
## Median : 34.00 Median : 8.75 Median : 0.8438 Median : 1.97193   
## Mean : 45.17 Mean :10.33 Mean : 1.5472 Mean : 3.07239   
## 3rd Qu.: 54.00 3rd Qu.:14.40 3rd Qu.: 1.8682 3rd Qu.: 3.80913   
## Max. :446.00 Max. :41.30 Max. :20.5613 Max. :27.03360   
## default   
## Min. :0.0000   
## 1st Qu.:0.0000   
## Median :0.0000   
## Mean :0.2667   
## 3rd Qu.:1.0000   
## Max. :1.0000

head(bankDetails)

## age ed employ address income debtinc creddebt othdebt default  
## 1 41 3 17 12 176 9.3 11.359392 5.008608 1  
## 2 27 1 10 6 31 17.3 1.362202 4.000798 0  
## 3 40 1 15 14 55 5.5 0.856075 2.168925 0  
## 4 41 1 15 14 120 2.9 2.658720 0.821280 0  
## 5 24 2 2 0 28 17.3 1.787436 3.056564 1  
## 6 41 2 5 5 25 10.2 0.392700 2.157300 0

Missing Data From summary : Observation : Data has no null values

Outliers :