Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

Programme	:	**BTech – ECE and ECM**	Semester	:	**Win 2022**
Course	:	**Essentials of Data Analytics Lab**	Code	:	**CSE3506**
Faculty	:	**Gobinath N**	Slot	:	**L51 + L52**

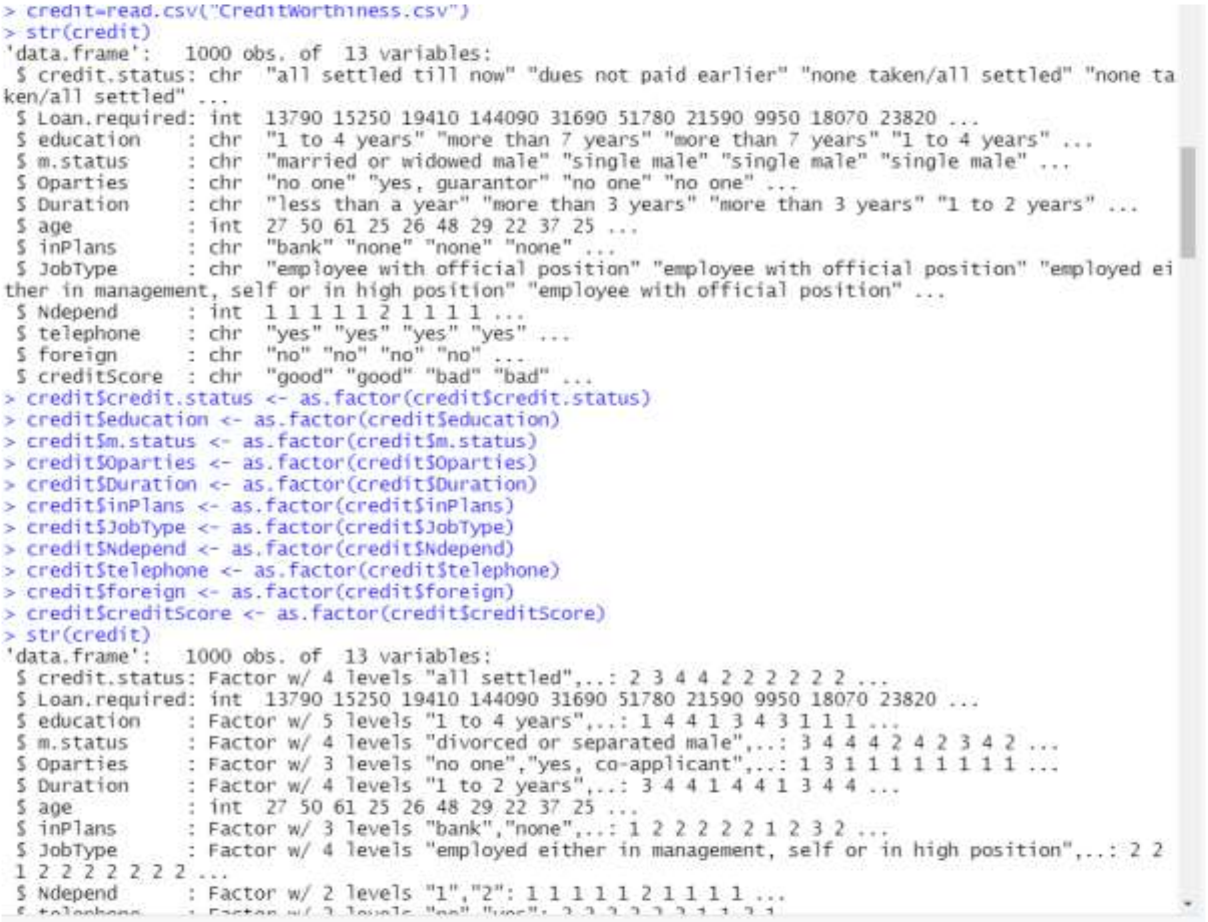
Ex_05/Classification Naive Bayes

Code:

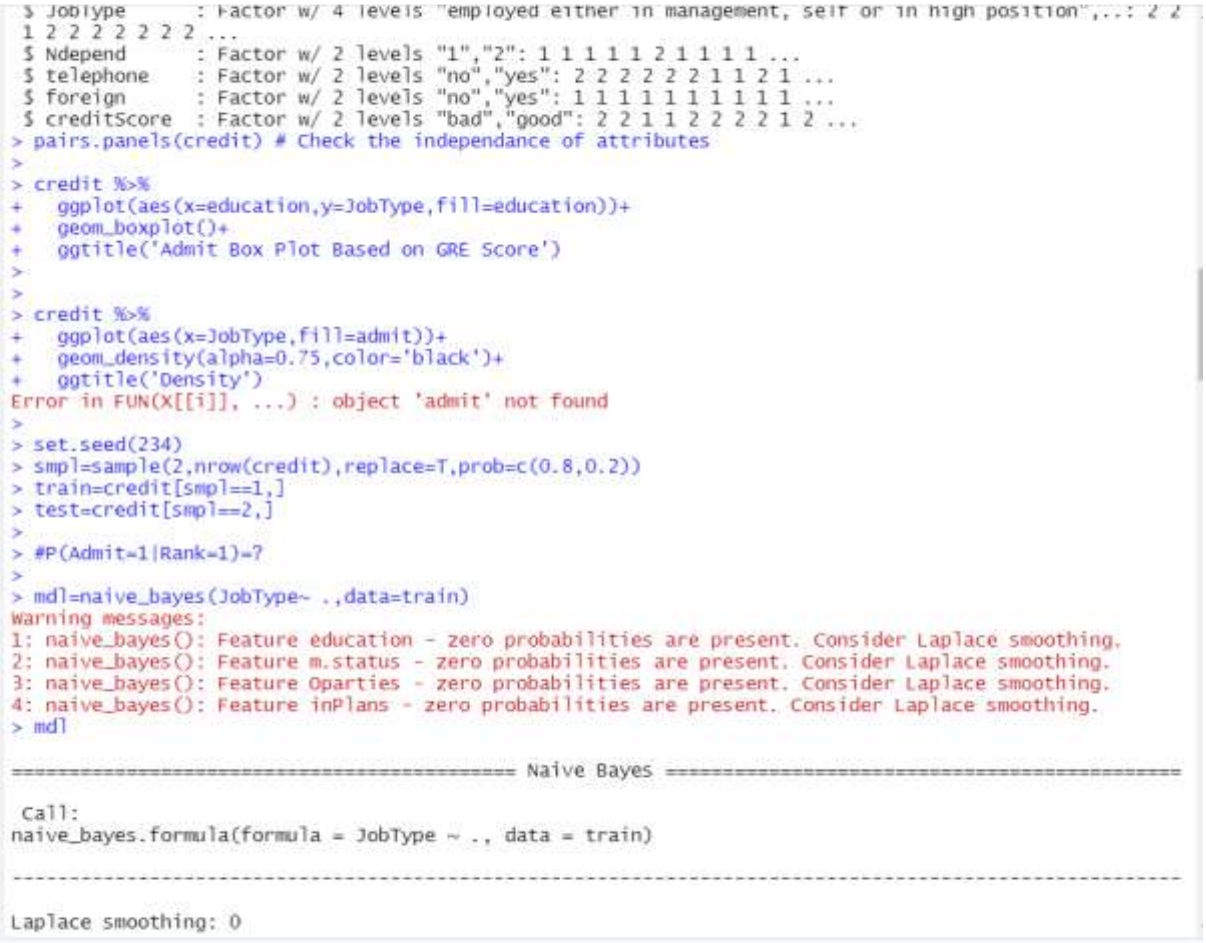
```
rm(list = ls())
setwd("C:\\Users\\Rituraj Anand\\Desktop\\Sem6\\CSE3506\\LAB\\Lab
5")
#install.packages('naivebayes')
#install.packages('psych')
library(naivebayes)
library(dplyr)
library(ggplot2)
library(psych)
credit=read.csv("CreditWorthiness.csv")
str(credit)
credit$credit.status <- as.factor(credit$credit.status)
credit$education <- as.factor(credit$education)
credit$m.status <- as.factor(credit$m.status)
credit$Oparties <- as.factor(credit$Oparties)
credit$Duration <- as.factor(credit$Duration)
```

```
credit$inPlans <- as.factor(credit$inPlans)
credit$JobType <- as.factor(credit$JobType)
credit$Ndepend <- as.factor(credit$Ndepend)
credit$telephone <- as.factor(credit$telephone)
credit$foreign <- as.factor(credit$foreign)
credit$creditScore <- as.factor(credit$creditScore)
str(credit)
pairs.panels(credit) # Check the independance of attributes

credit %>%
  ggplot(aes(x=education,y=JobType,fill=education))+
  geom_boxplot()+
  ggtitle('Admit Box Plot Based on GRE Score')



credit %>%
  ggplot(aes(x=JobType,fill=admit))+
  geom_density(alpha=0.75,color='black')+
  ggtitle('Density')

set.seed(234)
smpl=sample(2,nrow(credit),replace=T,prob=c(0.8,0.2))
train=credit[smpl==1,]
test=credit[smpl==2,]

#P(Admit=1|Rank=1)=?
```

```
mdl=naive_bayes(JobType~ .,data=train)

mdl

plot(mdl)


p=predict(mdl,train,type='prob')

head(cbind(p,train))


#To find the accuracy of prediction


p1=predict(mdl,train)

(tab1=table(p1,train$education))

1-sum(diag(tab1))/sum(tab1)
```

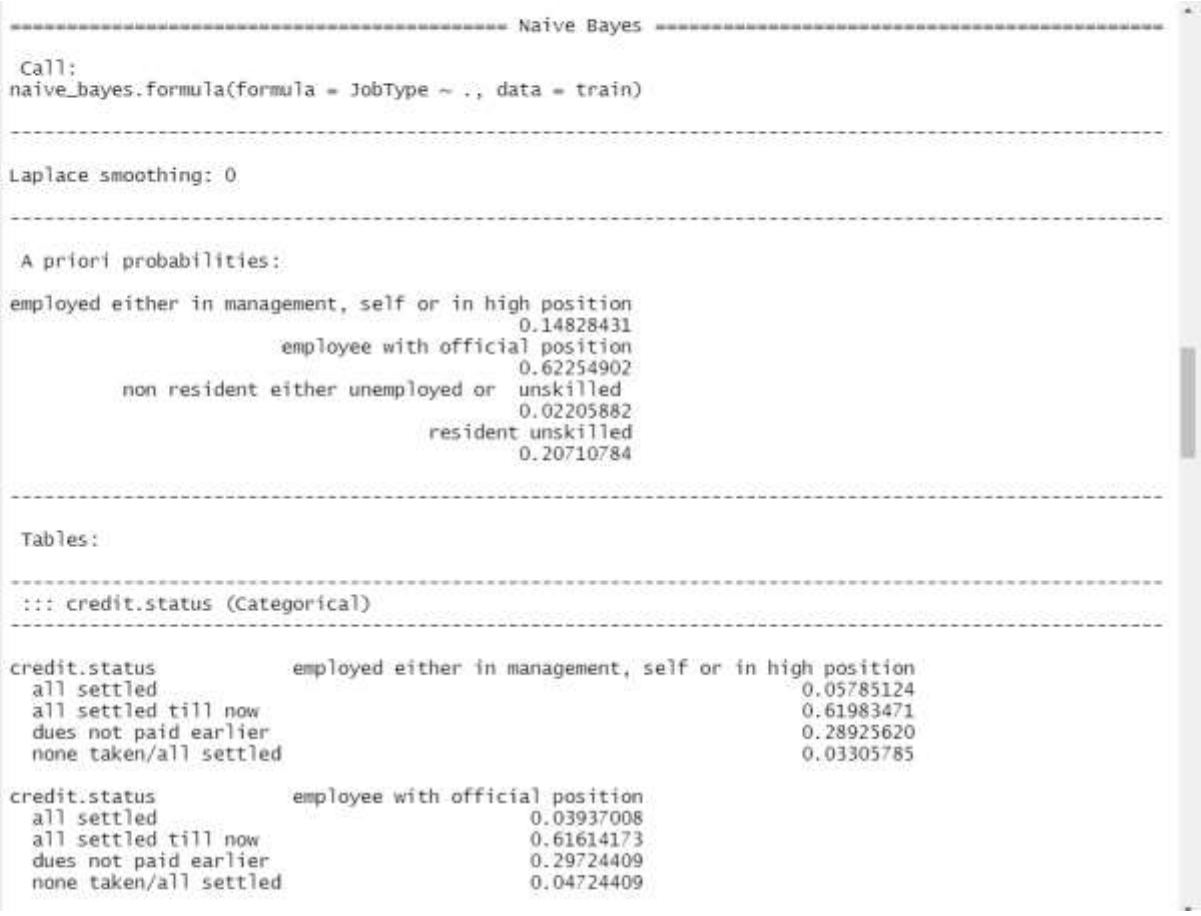


```
> credit=read.csv("CreditWorthiness.csv")
> str(credit)
'data.frame':   1000 obs. of  13 variables:
 $ credit.status: chr  "all settled till now" "dues not paid earlier" "none taken/all settled" "none ta
ken/all settled" ...
 $ Loan.required: int  13790 15250 19410 144090 31690 51780 21590 9950 18070 23820 ...
 $ education    : chr  "1 to 4 years" "more than 7 years" "more than 7 years" "1 to 4 years" ...
 $ m.status     : chr  "married or widowed male" "single male" "single male" "single male" ...
 $ Oparties     : chr  "no one" "yes, guarantor" "no one" "no one" ...
 $ Duration     : chr  "less than a year" "more than 3 years" "more than 3 years" "1 to 2 years" ...
 $ age          : int  27 50 61 25 26 48 29 22 37 25 ...
 $ inPlans      : chr  "bank" "none" "none" "none" ...
 $ JobType      : chr  "employee with official position" "employee with official position" "employed ei
ther in management, self or in high position" "employee with official position" ...
 $ Ndepend      : int  1 1 1 1 1 2 1 1 1 1 ...
 $ telephone    : chr  "yes" "yes" "yes" "yes" ...
 $ foreign      : chr  "no" "no" "no" "no" ...
 $ creditScore  : chr  "good" "good" "bad" "bad" ...
> credit$credit.status <- as.factor(credit$credit.status)
> credit$education <- as.factor(credit$education)
> credit$m.status <- as.factor(credit$m.status)
> credit$Oparties <- as.factor(credit$Oparties)
> credit$Duration <- as.factor(credit$Duration)
> credit$inPlans <- as.factor(credit$inPlans)
> credit$JobType <- as.factor(credit$JobType)
> credit$Ndepend <- as.factor(credit$Ndepend)
> credit$telephone <- as.factor(credit$telephone)
> credit$foreign <- as.factor(credit$foreign)
> credit$creditScore <- as.factor(credit$creditScore)
> str(credit)
'data.frame':   1000 obs. of  13 variables:
 $ credit.status: Factor w/ 4 levels "all settled",..: 2 3 4 4 2 2 2 2 2 2 ...
 $ Loan.required: int  13790 15250 19410 144090 31690 51780 21590 9950 18070 23820 ...
 $ education    : Factor w/ 5 levels "1 to 4 years",..: 1 4 4 1 3 4 3 1 1 1 ...
 $ m.status     : Factor w/ 4 levels "divorced or separated male",..: 3 4 4 4 2 4 2 3 4 2 ...
 $ Oparties     : Factor w/ 3 levels "no one","yes, co-applicant",..: 1 3 1 1 1 1 1 1 1 1 ...
 $ Duration     : Factor w/ 4 levels "1 to 2 years",..: 3 4 4 1 4 4 1 3 4 4 ...
 $ age          : int  27 50 61 25 26 48 29 22 37 25 ...
 $ inPlans      : Factor w/ 3 levels "bank","none",..: 1 2 2 2 2 2 1 2 3 2 ...
 $ JobType      : Factor w/ 4 levels "employed either in management, self or in high position",..: 2 2
 1 2 2 2 2 2 2 2 ...
 $ Ndepend      : Factor w/ 2 levels "1","2": 1 1 1 1 1 2 1 1 1 1 ...
 $ telephone    : Factor w/ 2 levels "no","yes": 2 2 2 2 2 2 1 1 2 1
```

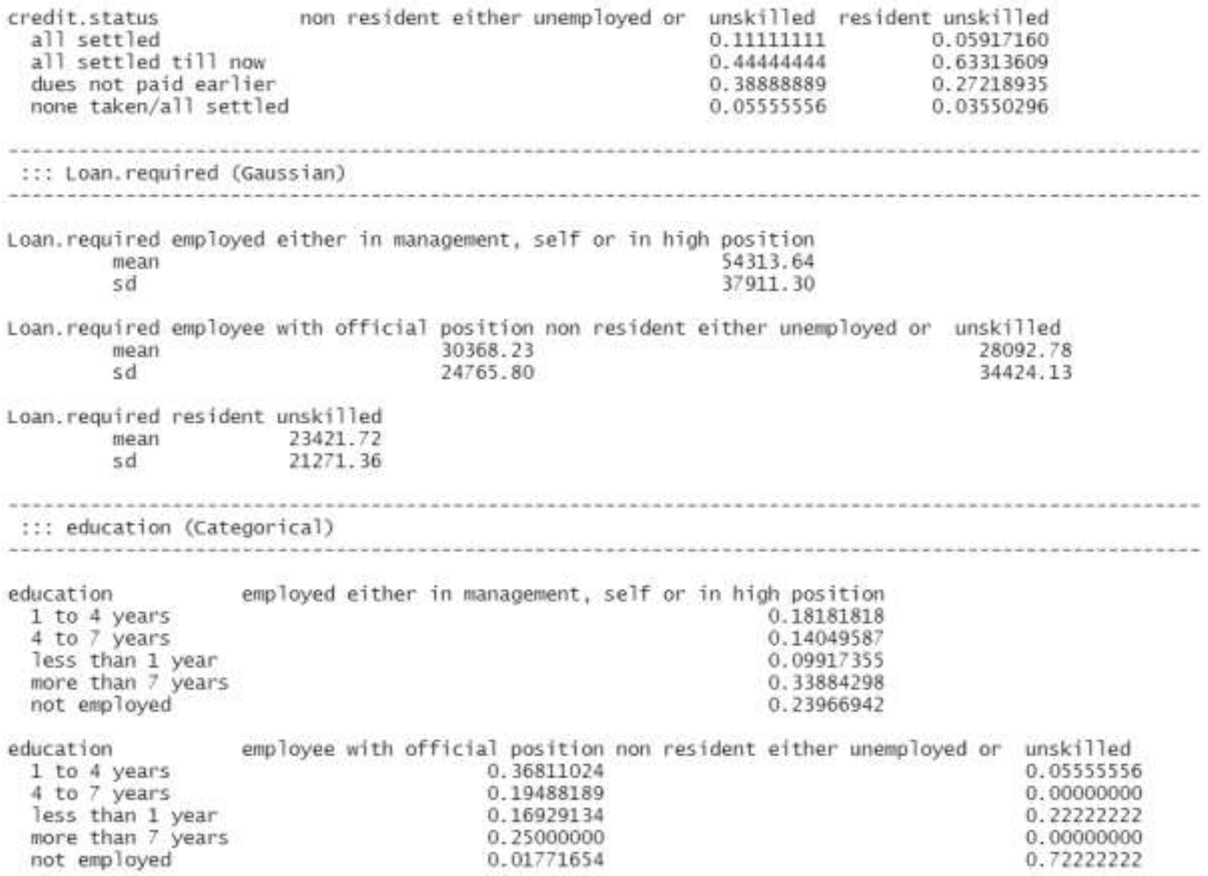


```
 $ JobType       : Factor w/ 4 levels "employed either in management, self or in high position",..: 2 2
 1 2 2 2 2 2 2 2 2 ...
 $ Ndepend       : Factor w/ 2 levels "1","2": 1 1 1 1 1 2 1 1 1 1 ...
 $ telephone     : Factor w/ 2 levels "no","yes": 2 2 2 2 2 2 1 1 2 1 ...
 $ foreign       : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
 $ creditScore   : Factor w/ 2 levels "bad","good": 2 2 1 1 2 2 2 2 1 2 ...
> pairs.panels(credit) # Check the independance of attributes
>
> credit %>%
+   ggplot(aes(x=education,y=JobType,fill=education))+
+   geom_boxplot()+
+   ggtitle('Admit Box Plot Based on GRE Score')
>
>
> credit %>%
+   ggplot(aes(x=JobType,fill=admit))+
+   geom_density(alpha=0.75,color='black')+
+   ggtitle('Density')
Error in FUN(X[[i]], ...) : object 'admit' not found
>
> set.seed(234)
> smpl=sample(2,nrow(credit),replace=T,prob=c(0.8,0.2))
> train=credit[smpl==1,]
> test=credit[smpl==2,]
>
> #P(Admit=1|Rank=1)=?
>
> mdl=naive_bayes(JobType~ .,data=train)
Warning messages:
1: naive_bayes(): Feature education - zero probabilities are present. Consider Laplace smoothing.
2: naive_bayes(): Feature m.status - zero probabilities are present. Consider Laplace smoothing.
3: naive_bayes(): Feature Oparties - zero probabilities are present. Consider Laplace smoothing.
4: naive_bayes(): Feature inPlans - zero probabilities are present. Consider Laplace smoothing.
> mdl

================================================ Naive Bayes =============================================

 Call:
naive_bayes.formula(formula = JobType ~ ., data = train)

----------------------------------------------------------------------------------------------------------

Laplace smoothing: 0
```

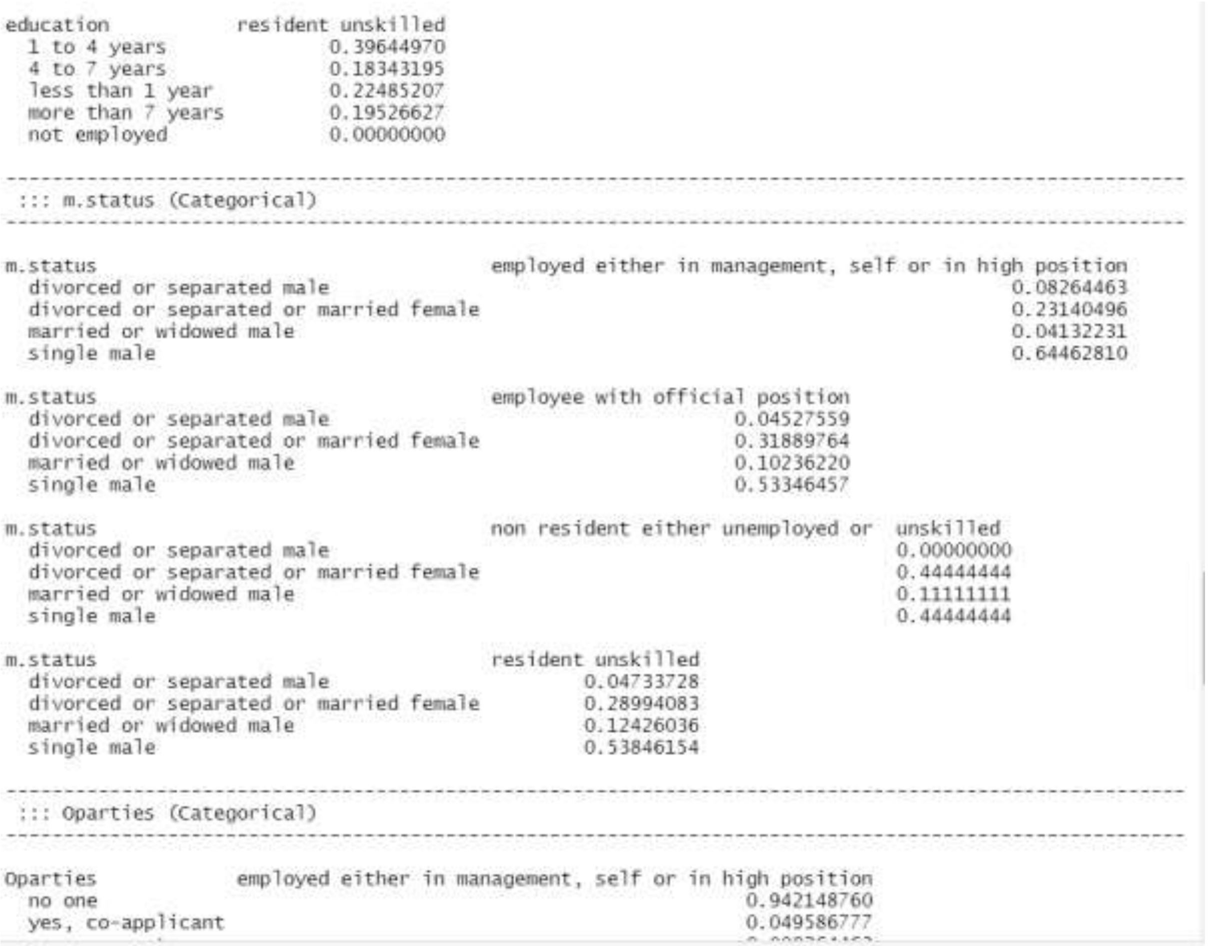
Naïve Baye’s



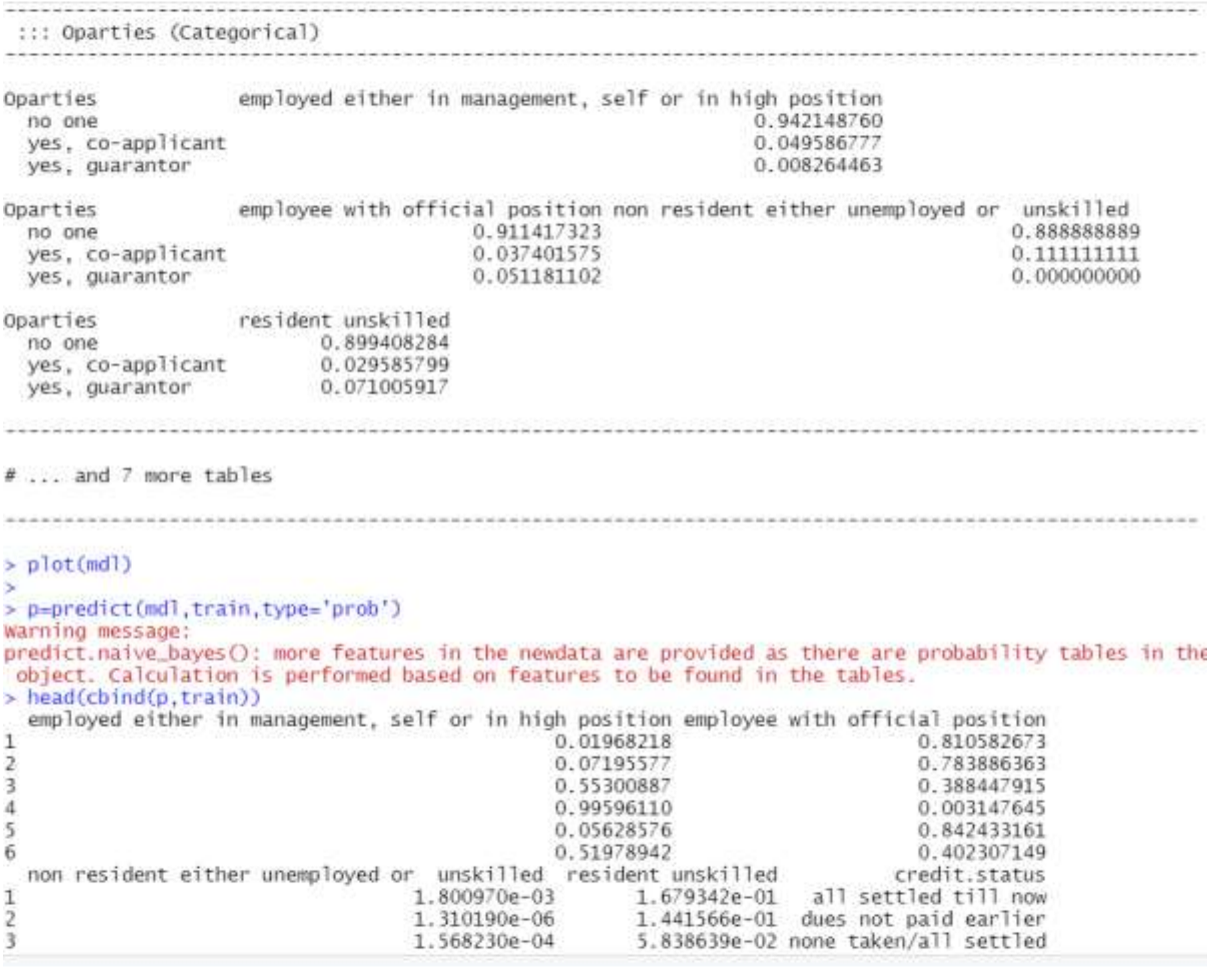
```
============================================ Naive Bayes ============================================

 Call:
naive_bayes.formula(formula = JobType ~ ., data = train)

-----------------------------------------------------------------------------------------------------

Laplace smoothing: 0

-----------------------------------------------------------------------------------------------------

 A priori probabilities:

employed either in management, self or in high position
                                             0.14828431
                        employee with official position
                                             0.62254902
          non resident either unemployed or  unskilled
                                             0.02205882
                                     resident unskilled
                                             0.20710784

-----------------------------------------------------------------------------------------------------

 Tables:

-----------------------------------------------------------------------------------------------------
 ::: credit.status (Categorical)
-----------------------------------------------------------------------------------------------------

credit.status            employed either in management, self or in high position
  all settled                                                         0.05785124
  all settled till now                                                0.61983471
  dues not paid earlier                                               0.28925620
  none taken/all settled                                              0.03305785

credit.status            employee with official position
  all settled                                 0.03937008
  all settled till now                        0.61614173
  dues not paid earlier                       0.29724409
  none taken/all settled                      0.04724409
```



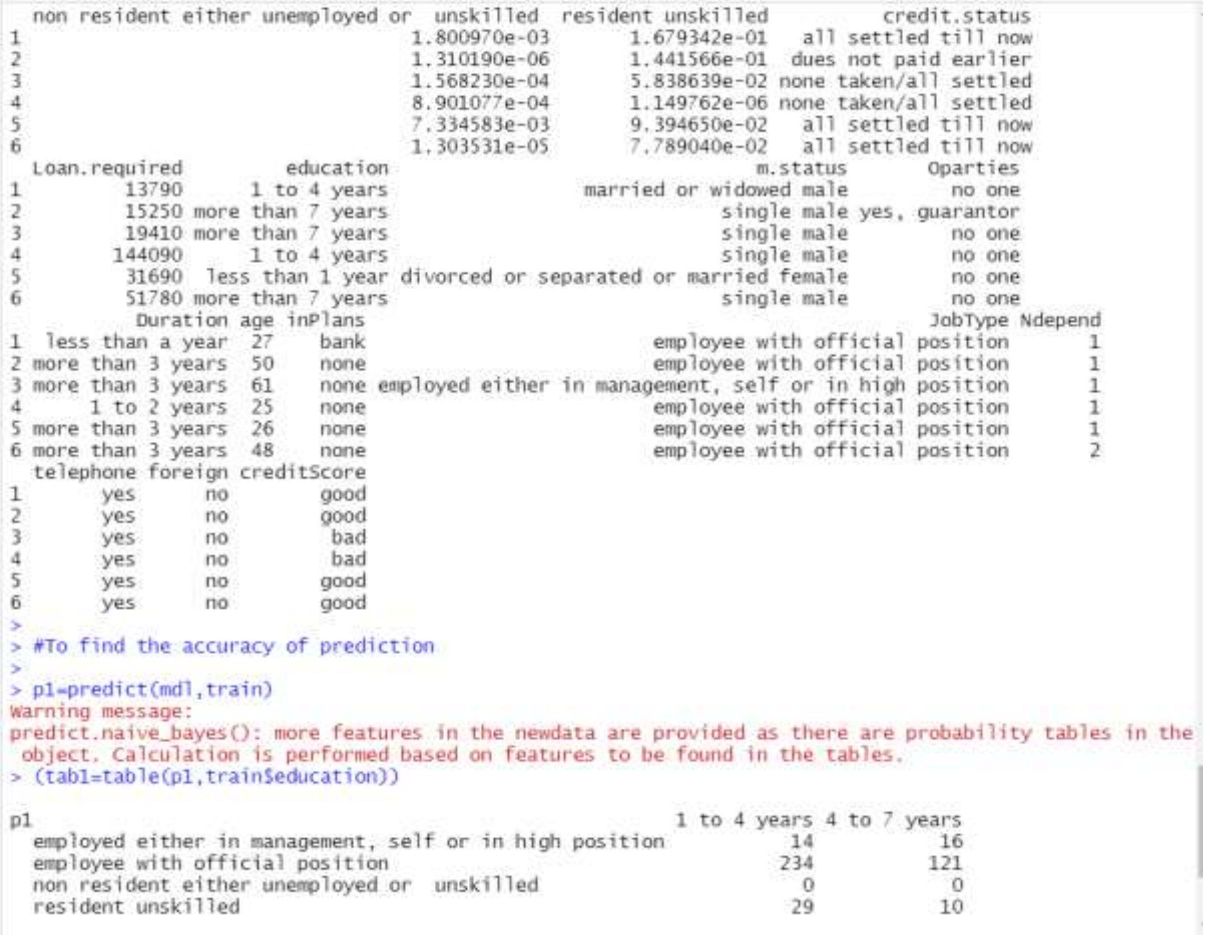
```
credit.status           non resident either unemployed or  unskilled  resident unskilled
  all settled                                              0.11111111         0.05917160
  all settled till now                                     0.44444444         0.63313609
  dues not paid earlier                                    0.38888889         0.27218935
  none taken/all settled                                   0.05555556         0.03550296

---------------------------------------------------------------------------------------------
 ::: Loan.required (Gaussian)
---------------------------------------------------------------------------------------------

Loan.required employed either in management, self or in high position
        mean                                                 54313.64
        sd                                                   37911.30

Loan.required employee with official position non resident either unemployed or  unskilled
        mean                        30368.23                                        28092.78
        sd                          24765.80                                        34424.13

Loan.required resident unskilled
        mean            23421.72
        sd              21271.36

---------------------------------------------------------------------------------------------
 ::: education (Categorical)
---------------------------------------------------------------------------------------------

education          employed either in management, self or in high position
  1 to 4 years                                                  0.18181818
  4 to 7 years                                                  0.14049587
  less than 1 year                                              0.09917355
  more than 7 years                                             0.33884298
  not employed                                                  0.23966942

education          employee with official position non resident either unemployed or  unskilled
  1 to 4 years                       0.36811024                                      0.05555556
  4 to 7 years                       0.19488189                                      0.00000000
  less than 1 year                   0.16929134                                      0.22222222
  more than 7 years                  0.25000000                                      0.00000000
  not employed                       0.01771654                                      0.72222222
```



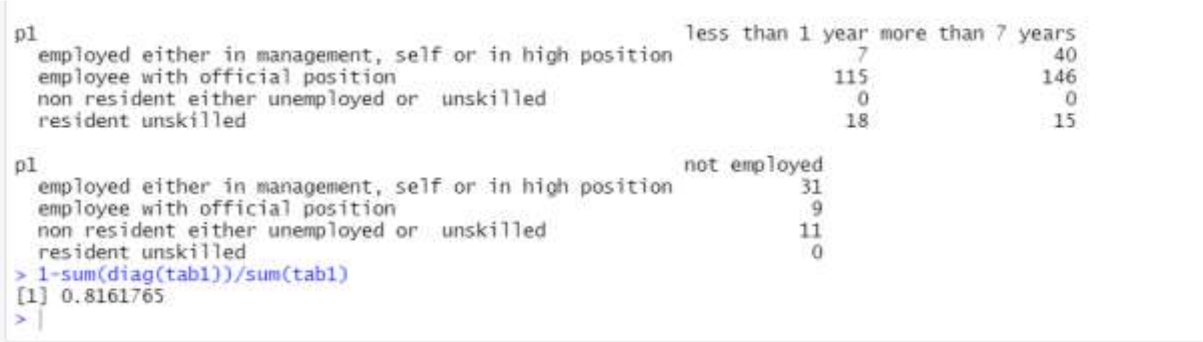
```
education            resident unskilled
  1 to 4 years               0.39644970
  4 to 7 years               0.18343195
  less than 1 year           0.22485207
  more than 7 years          0.19526627
  not employed               0.00000000

------------------------------------------------------------------------------------------------------
 ::: m.status (Categorical)
------------------------------------------------------------------------------------------------------

m.status                                    employed either in management, self or in high position
  divorced or separated male                                                             0.08264463
  divorced or separated or married female                                                0.23140496
  married or widowed male                                                                0.04132231
  single male                                                                            0.64462810

m.status                                    employee with official position
  divorced or separated male                                      0.04527559
  divorced or separated or married female                         0.31889764
  married or widowed male                                         0.10236220
  single male                                                     0.53346457

m.status                                    non resident either unemployed or  unskilled
  divorced or separated male                                                   0.00000000
  divorced or separated or married female                                      0.44444444
  married or widowed male                                                      0.11111111
  single male                                                                  0.44444444

m.status                                    resident unskilled
  divorced or separated male                        0.04733728
  divorced or separated or married female           0.28994083
  married or widowed male                           0.12426036
  single male                                       0.53846154

------------------------------------------------------------------------------------------------------
 ::: Oparties (Categorical)
------------------------------------------------------------------------------------------------------

Oparties             employed either in management, self or in high position
  no one                                                      0.942148760
  yes, co-applicant                                           0.049586777
```



```
-----------------------------------------------------------------------------------------------------
 ::: Oparties (Categorical)
-----------------------------------------------------------------------------------------------------

Oparties            employed either in management, self or in high position
  no one                                                        0.942148760
  yes, co-applicant                                             0.049586777
  yes, guarantor                                                0.008264463

Oparties            employee with official position non resident either unemployed or  unskilled
  no one                                0.911417323                                   0.888888889
  yes, co-applicant                     0.037401575                                   0.111111111
  yes, guarantor                        0.051181102                                   0.000000000

Oparties            resident unskilled
  no one                   0.899408284
  yes, co-applicant        0.029585799
  yes, guarantor           0.071005917

-----------------------------------------------------------------------------------------------------

# ... and 7 more tables

-----------------------------------------------------------------------------------------------------

> plot(mdl)
> 
> p=predict(mdl,train,type='prob')
Warning message:
predict.naive_bayes(): more features in the newdata are provided as there are probability tables in the
 object. Calculation is performed based on features to be found in the tables.
> head(cbind(p,train))
  employed either in management, self or in high position employee with official position
1                                              0.01968218                     0.810582673
2                                              0.07195577                     0.783886363
3                                              0.55300887                     0.388447915
4                                              0.99596110                     0.003147645
5                                              0.05628576                     0.842433161
6                                              0.51978942                     0.402307149
  non resident either unemployed or  unskilled  resident unskilled          credit.status
1                                 1.800970e-03        1.679342e-01   all settled till now
2                                 1.310190e-06        1.441566e-01  dues not paid earlier
3                                 1.568230e-04        5.838639e-02 none taken/all settled
```

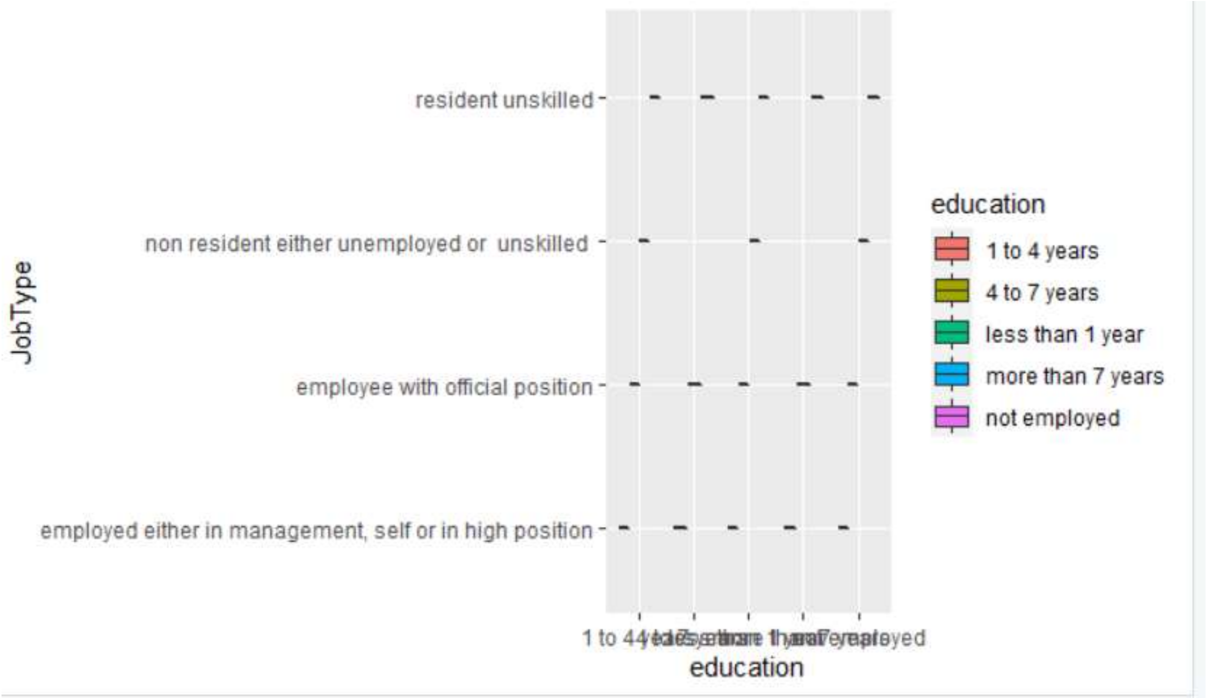


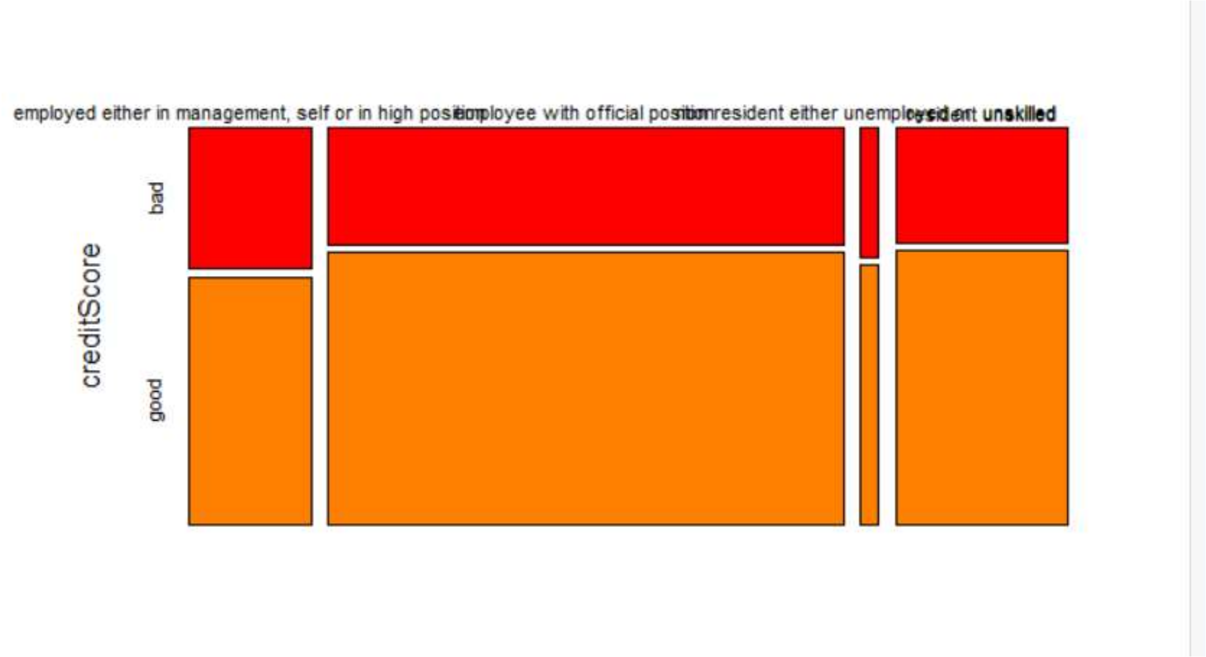
```
  non resident either unemployed or  unskilled  resident unskilled         credit.status
1                                  1.800970e-03       1.679342e-01  all settled till now
2                                  1.310190e-06       1.441566e-01  dues not paid earlier
3                                  1.568230e-04       5.838639e-02 none taken/all settled
4                                  8.901077e-04       1.149762e-06 none taken/all settled
5                                  7.334583e-03       9.394650e-02  all settled till now
6                                  1.303531e-05       7.789040e-02  all settled till now
  Loan.required         education                             m.status        Oparties
1         13790      1 to 4 years             married or widowed male          no one
2         15250 more than 7 years                          single male yes, guarantor
3         19410 more than 7 years                          single male          no one
4        144090      1 to 4 years                          single male          no one
5         31690  less than 1 year divorced or separated or married female          no one
6         51780 more than 7 years                          single male          no one
          Duration age inPlans                                        JobType Ndepend
1  less than a year  27    bank                    employee with official position       1
2 more than 3 years  50    none                    employee with official position       1
3 more than 3 years  61    none employed either in management, self or in high position       1
4      1 to 2 years  25    none                    employee with official position       1
5 more than 3 years  26    none                    employee with official position       1
6 more than 3 years  48    none                    employee with official position       2
  telephone foreign creditScore
1       yes      no        good
2       yes      no        good
3       yes      no         bad
4       yes      no         bad
5       yes      no        good
6       yes      no        good
>
> #To find the accuracy of prediction
>
> p1=predict(mdl,train)
Warning message:
predict.naive_bayes(): more features in the newdata are provided as there are probability tables in the
 object. Calculation is performed based on features to be found in the tables.
> (tab1=table(p1,train$education))

p1                                                  1 to 4 years 4 to 7 years
  employed either in management, self or in high position           14           16
  employee with official position                                   234          121
  non resident either unemployed or  unskilled                         0            0
  resident unskilled                                                  29           10
```

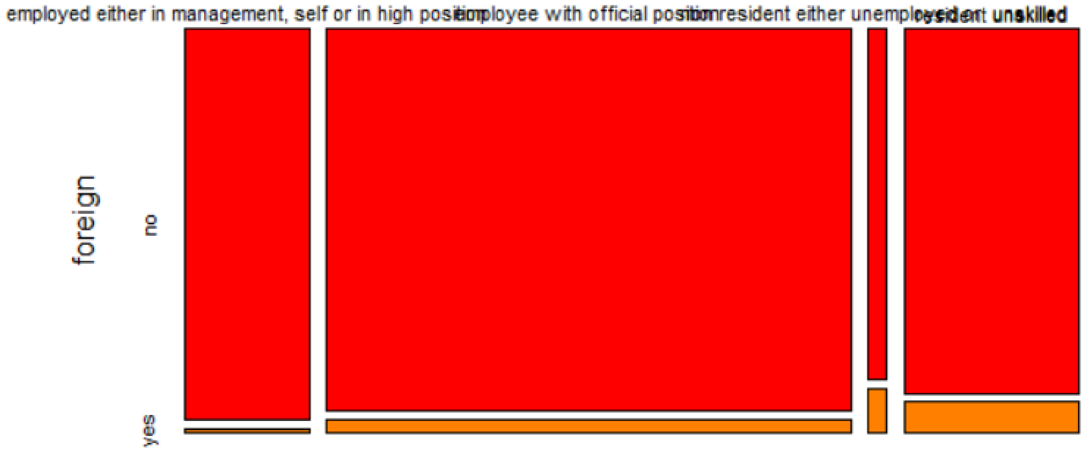


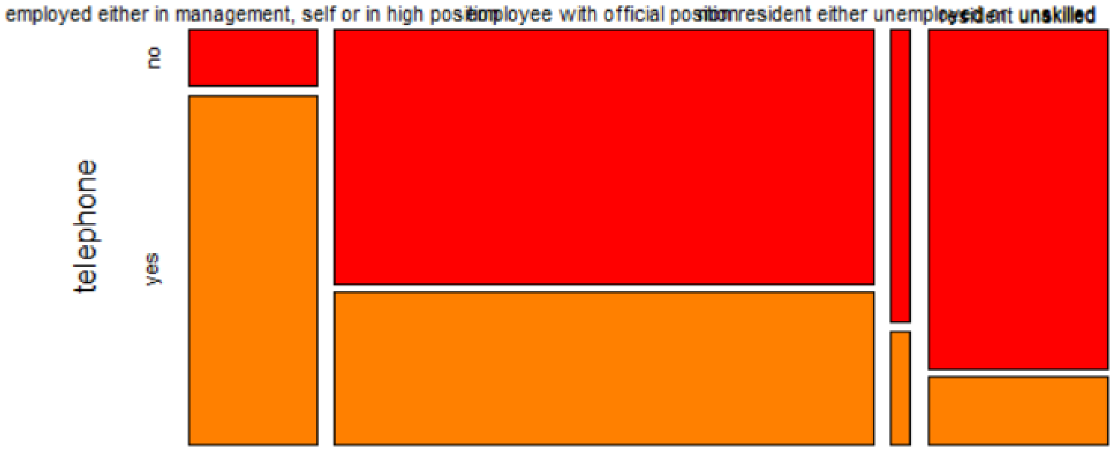
```
p1                                                  less than 1 year more than 7 years
  employed either in management, self or in high position                7                40
  employee with official position                                      115               146
  non resident either unemployed or  unskilled                            0                 0
  resident unskilled                                                     18                15

p1                                                  not employed
  employed either in management, self or in high position           31
  employee with official position                                    9
  non resident either unemployed or  unskilled                       11
  resident unskilled                                                 0
> 1-sum(diag(tab1))/sum(tab1)
[1] 0.8161765
> |
```

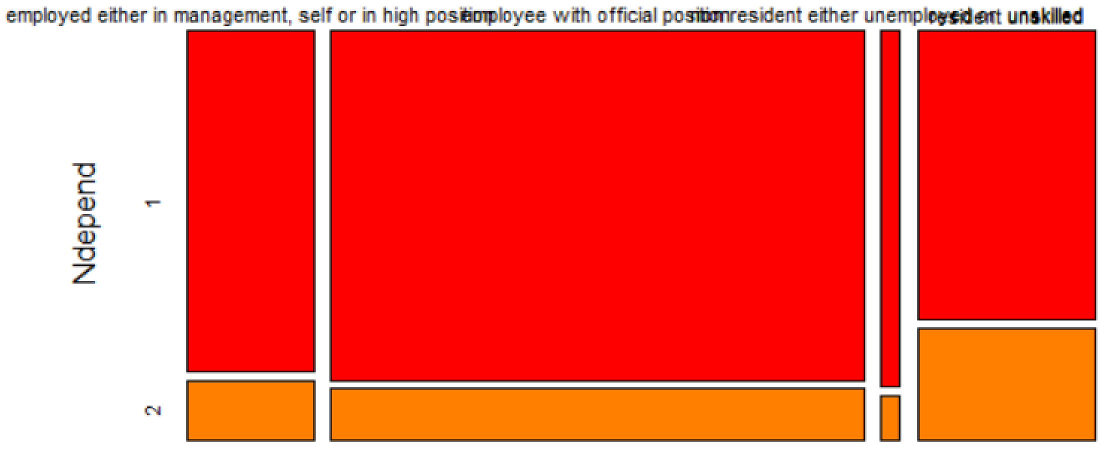
Box Plot

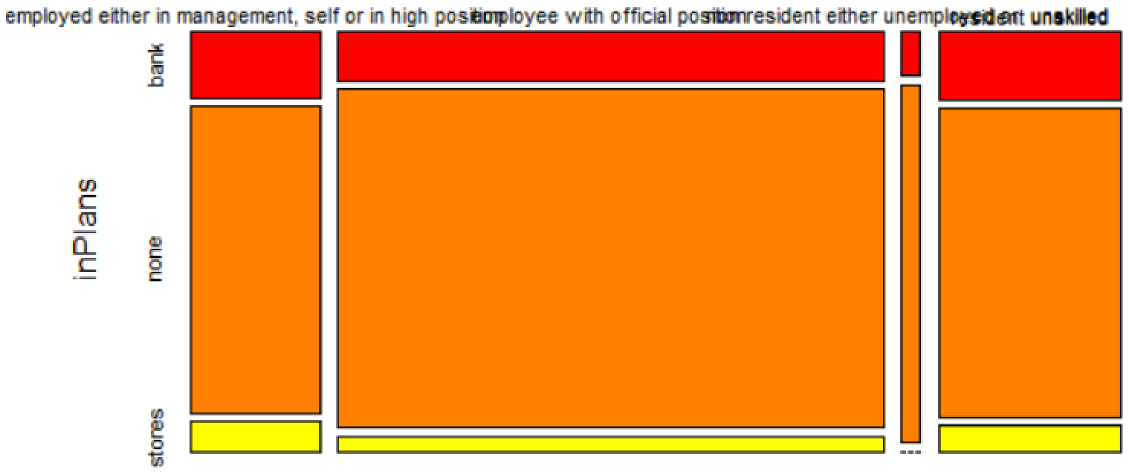

resident unskilled
non resident either unemployed or unskilled
employee with official position
employed either in management, self or in high position
JobType
education
1 to 4 years
4 to 7 years
less than 1 year
more than 7 years
not employed
education

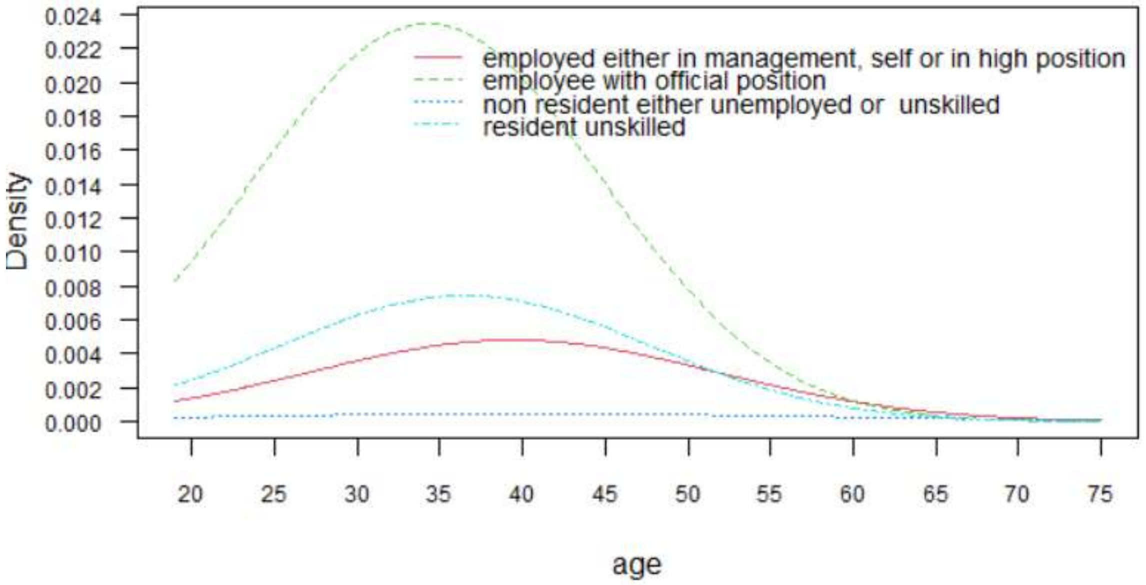

employed either in management, self or in high position
employee with official position
non resident either unemployed or unskilled
resident unskilled
creditScore
bad
good

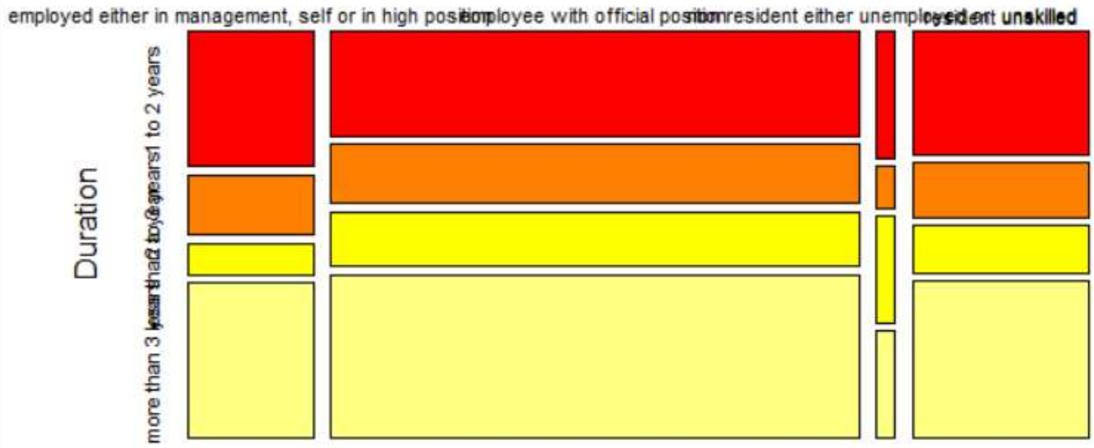

employed either in management, self or in high position
employee with official position
non resident either unemployed or unskilled
resident unskilled
foreign
no
yes

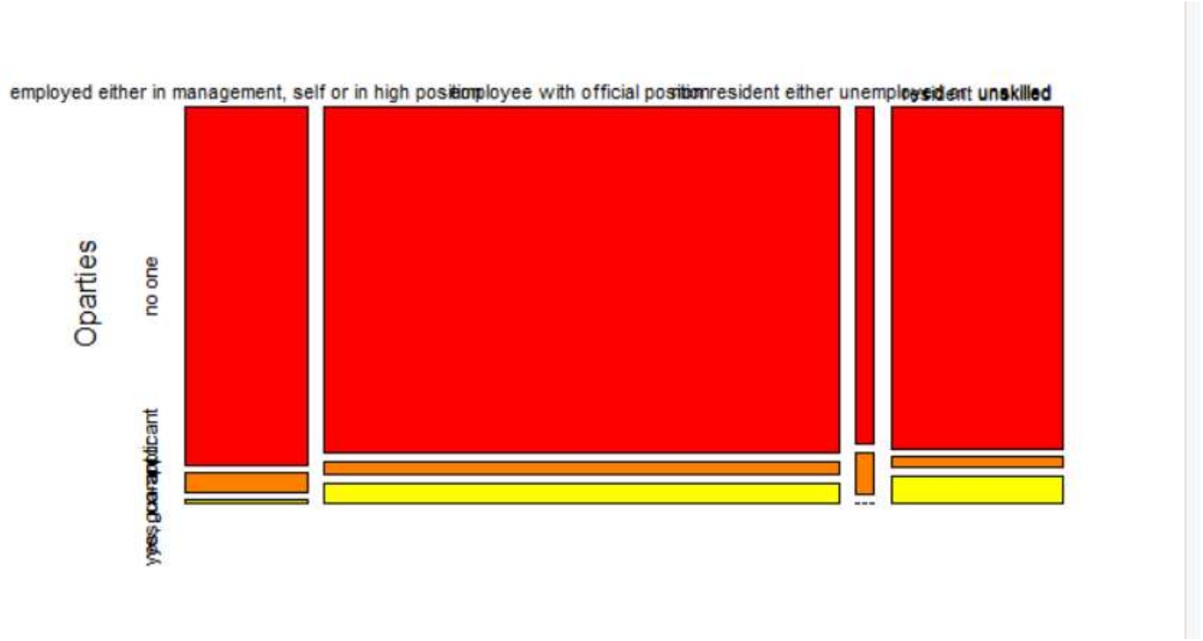

employed either in management, self or in high position
employee with official position
non resident either unemployed or unskilled
resident unskilled
telephone
no
yes

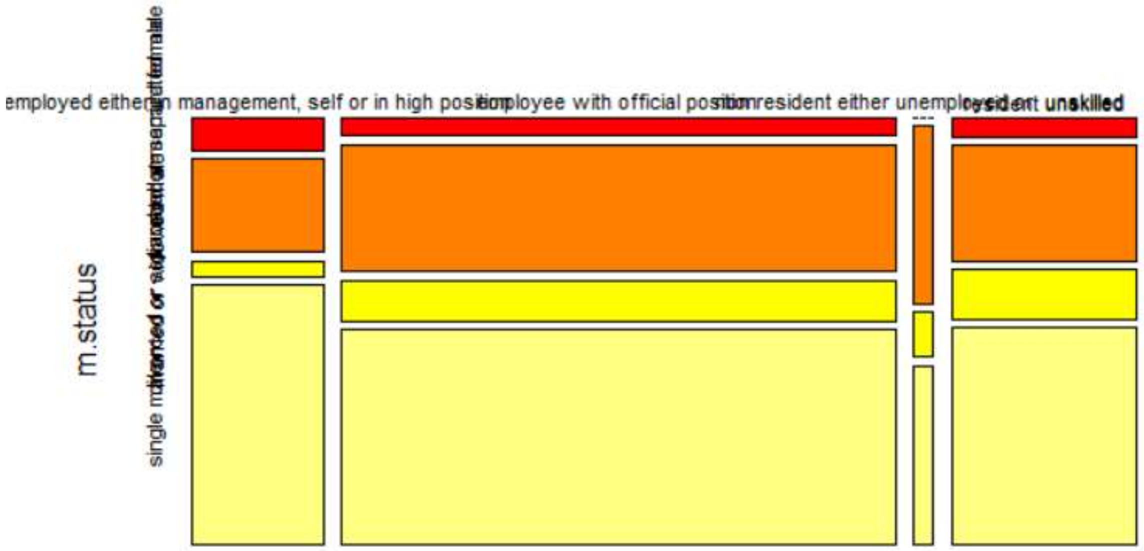

employed either in management, self or in high position
employee with official position
non resident either unemployed or unskilled
resident unskilled
Ndepend
1
2

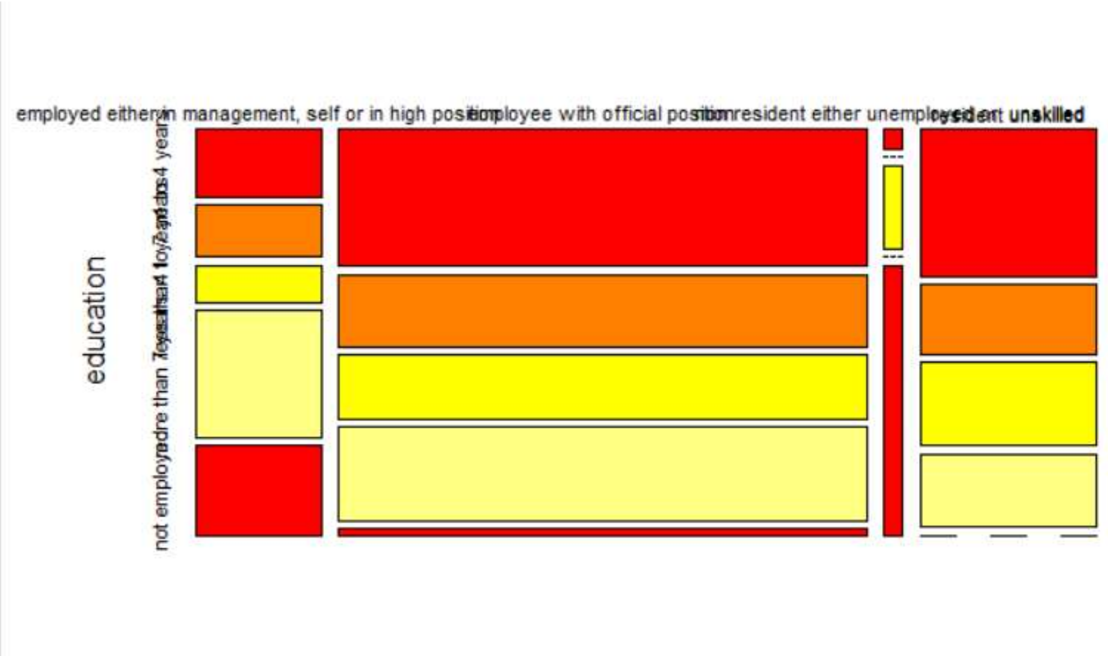

employed either in management, self or in high position
employee with official position
non resident either unemployed or unskilled
resident unskilled
inPlans
bank
none
stores

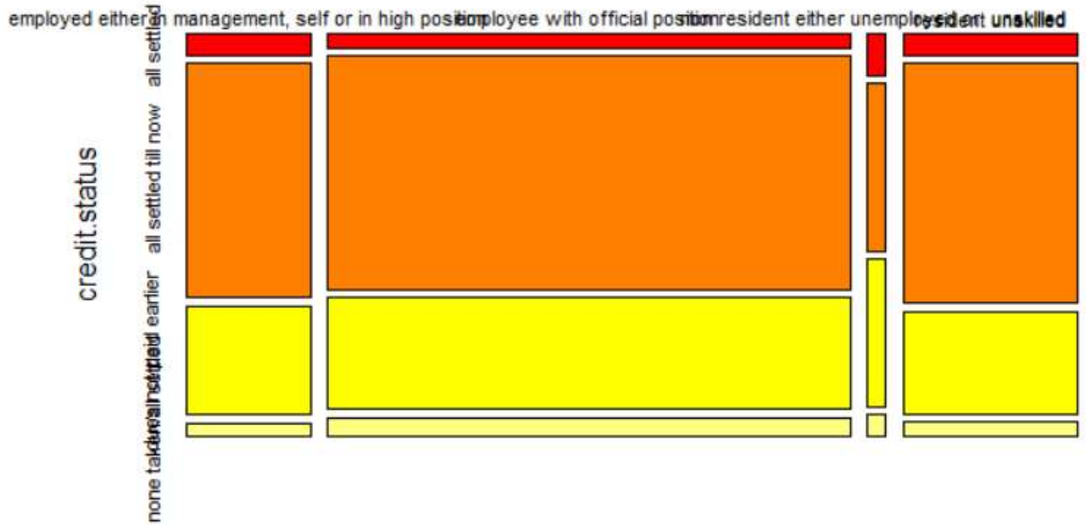

Density
0.024
0.022
0.020
0.018
0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
20
25
30
35
40
45
50
55
60
65
70
75
age
employed either in management, self or in high position
employee with official position
non resident either unemployed or unskilled
resident unskilled


employed either in management, self or in high position
employee with official position
non resident either unemployed or unskilled
resident unskilled
Duration
1 to 2 years
2 to 3 years
less than a year
more than 3 years


employed either in management, self or in high position
employee with official position
non resident either unemployed
resident unskilled
Oparties
no one
applicant
yes


employed either in management, self or in high position
employee with official position
non resident either unemployed
resident unskilled
m.status
single m
female





Result and Inference:

Hence, we saw how naïve baye’s theorem is used for predication using different plots applied on Credit worthiness set based on education and JobType.