TECNOLÓGICO NACIONAL DE MÉXICO

INSTITUTO TECNOLÓGICO DE TIJUANA

Subdirección Académica

Departamento de sistemas y computación

Semestre:
Agosto - Diciembre 2021

Materia:
Minería de Datos

Nombre del trabajo:
Práctica Evaluatoria

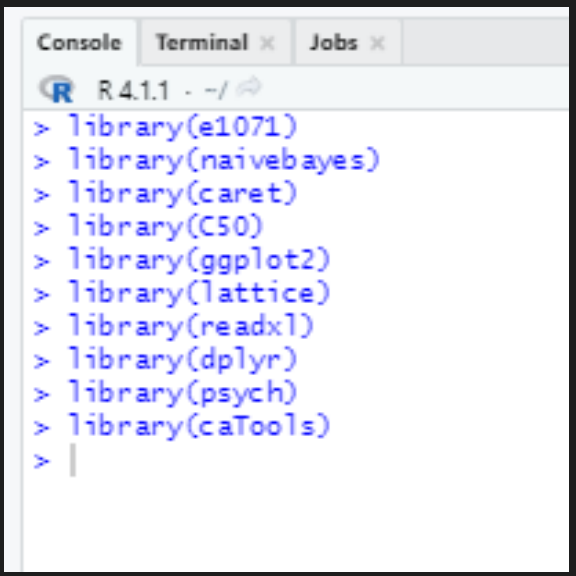
UNIDAD A EVALUAR:
Unidad III

NOMBRE Y NÚMERO DE CONTROL DEL ALUMNO:
Victor Fabian Hernandez Lamarque | 16212536
Cristopher Arael Garcia Torres | 17210892

NOMBRE DEL MAESTRO (A):
JOSE CHRISTIAN ROMERO

```r
#Usamos nuestras bibliotecas ya instaladas con ctrl + enter

Install.packages('e1071')
library(e1071)
Install.packages('naivebayes')
library(naivebayes)
Install.packages('caret')
library(caret)
Install.packages('C50')
library(C50)
Install.packages('ggplot2')
library(ggplot2)
Install.packages('lattice')
library(lattice)
Install.packages('readxl')
library(readxl)
Install.packages('dplyr')
library(dplyr)
Install.packages('psych')
library(psych)
Install.packages('caTools')
library(caTools)
```



```r
#importamos nuestro dataset
dataset<-read.csv("Social_Network_Ads.csv")
```

```
> dataset
     User.ID Gender Age EstimatedSalary Purchased
1   15624510   Male  19            19000         0
2   15810944   Male  35            20000         0
3   15668575 Female  26            43000         0
4   15603246 Female  27            57000         0
5   15804002   Male  19            76000         0
6   15728773   Male  27            58000         0
7   15598044 Female  27            84000         0
8   15694829 Female  32           150000         1
9   15600575   Male  25            33000         0
10  15727311 Female  35            65000         0
11  15570769 Female  26            80000         0
12  15606274 Female  26            52000         0
13  15746139   Male  20            86000         0
14  15704987   Male  32            18000         0
15  15628972   Male  18            82000         0
```

```r
#le indicamos como queremos nuestro dataset

dataset = dataset[3:5]

dataset$Purchased=factor(dataset$Purchased,levels = c(0,1))

set.seed(123)

split=sample.split(dataset$Purchased, SplitRatio = 0.45)

training_set=subset(dataset,split==TRUE)
test_set=subset(dataset,split==FALSE)



training_set[-2]=scale(training_set[-2])
test_set[-2]=scale(test_set[-2])
```

```
test_set          220 obs. of 3 variables
   $ Age            : num  -0.3 -1.07 -1.85 -1.07 -0.59 ...
   $ EstimatedSalary: num  -1.52 -0.393 0.186 0.43 2.441 ...
   $ Purchased      : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 1 1 1 ...
training_set      180 obs. of 3 variables
   $ Age            : num  -1.698 -1.042 -0.948 -0.761 0.832 ...
   $ EstimatedSalary: num  -1.417 -0.744 -0.324 0.293 -1.304 ...
   $ Purchased      : Factor w/ 2 levels "0","1": 1 1 1 1 2 2 2 1 1 1 ...
Values
  split           logi [1:400] TRUE FALSE TRUE FALSE FALSE TRUE ...
```

```
#utilizamos la fórmula de naive bayes
classifier=naive_bayes(formula=Purchased ~ . ,
                       data=training_set,
                       type='C-classification',
                       kernel='linear')

y_pred=predict(classifier,newdata=test_set[-3])
y_pred
```

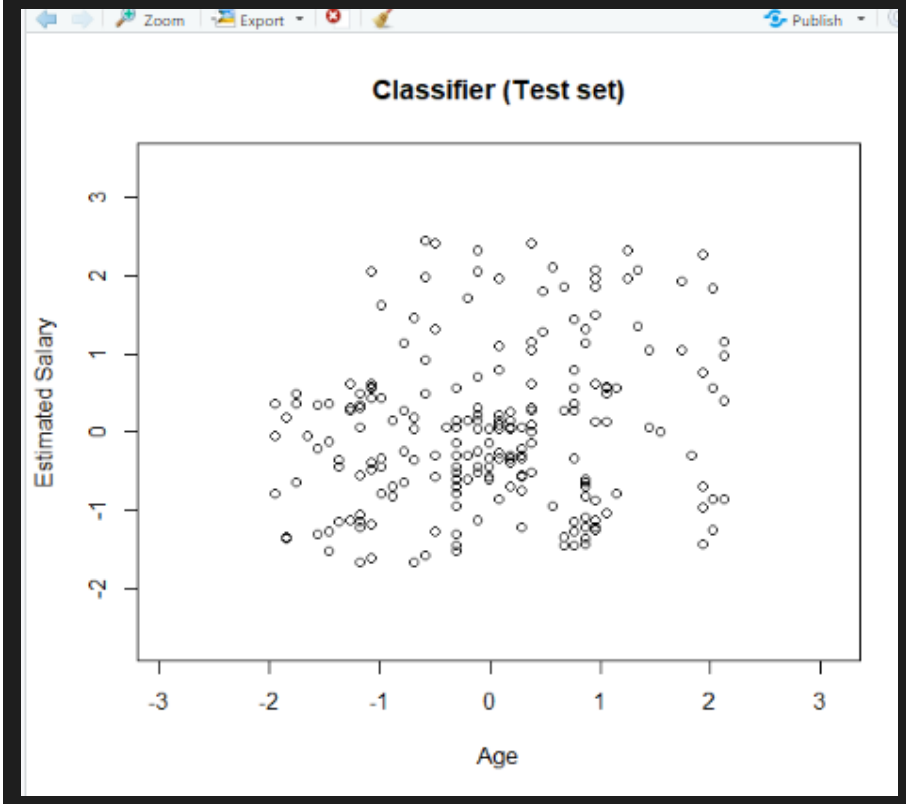
```
> y_pred=predict(classifier,newdata=test_set[-3])
> y_pred
  [1] 0 0 0 0 1 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0
 [67] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 0 1 1 0 0 1 1 0 0 1 1 0 0 1
[133] 1 0 1 1 1 0 0 1 1 1 0 0 1 1 0 1 1 1 0 0 1 1 0 0 1 1 1 0 0 1 1 1 1 1 1 0 1 1 0 0 0 0 0 0 1 0 1 1 0 0 1 0 1 1 1 0 1 0 0 1 0 0 0 1 0 0
[199] 1 1 0 1 1 1 1 0 1 0 1 1 0 1 1 1 0 1 1 1 0 1
Levels: 0 1
>
```

```
cm=table(test_set[, 3],y_pred)
cm
```

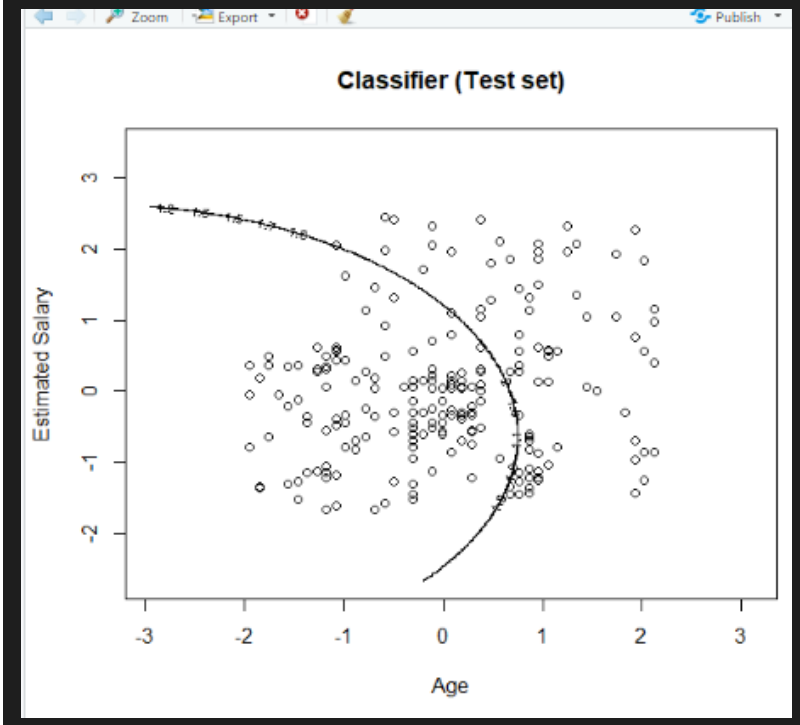
```
> cm=table(test_set[, 3],y_pred)
> cm
   y_pred
      0   1
  0 129  12
  1  12  67
>
```

```
set = test_set
X1 = seq(min(set[, 1]) - 1, max(set[, 1]) + 1, by = 0.01)
X2 = seq(min(set[, 2]) - 1, max(set[, 2]) + 1, by = 0.01)
grid_set = expand.grid(X1, X2)
#guardamos en nuestra variable grid_set los nombres de las columnas
colnames(grid_set) = c('Age', 'EstimatedSalary')
```

```
#visualizamos la gráfica por la edad y el salario estimado
y_grid = predict(classifier, newdata = grid_set)
plot(set[, -3], main = 'Classifier (Test set)',
     xlab = 'Age', ylab = 'Estimated Salary',
     xlim = range(X1), ylim = range(X2))
```



```
#le agregamos la parábola a nuestra grafica
contour(X1, X2, matrix(as.numeric(y_grid), length(X1), length(X2)), add
=TRUE)
```




```
#le asignamos colores a los puntos de dispersión y a otra la gráfica
para una mayor mejor en cuanto a la visualización
points(grid_set, pch = '.', col = ifelse(y_grid == 1, 'pink','red'))
points(set, pch = 21, bg = ifelse(set[, 3] == 1, 'yellow', 'green'))
```