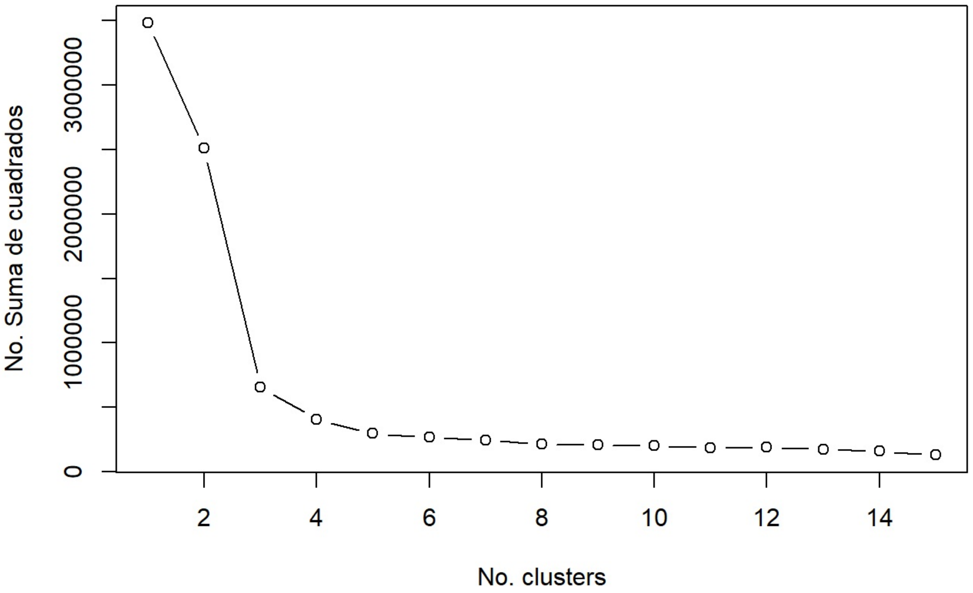
Algoritmos

Cristopher Barrios, Elean Rivas, Angel Higueros, Mariana David

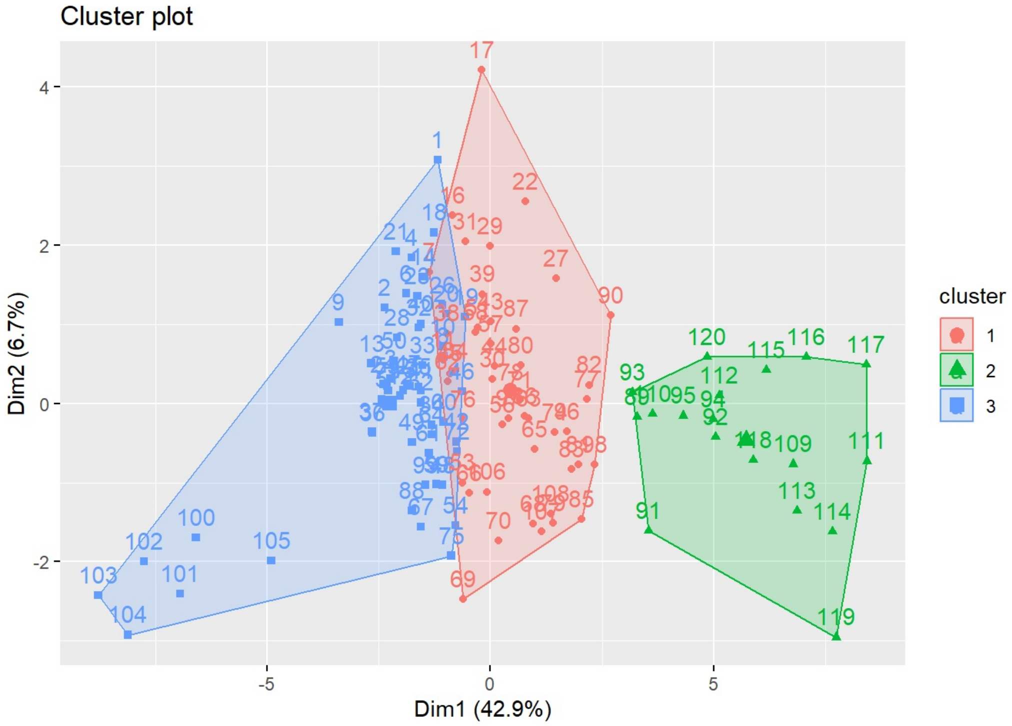
16/2/2023

Resultados del proyecto

Grafica de codos de divorcios por departamento por mes



Numero de Clusters (Divorcios por departamento por mes)



Aplicacion de clusters a Data Inicial y Separacion de grupos

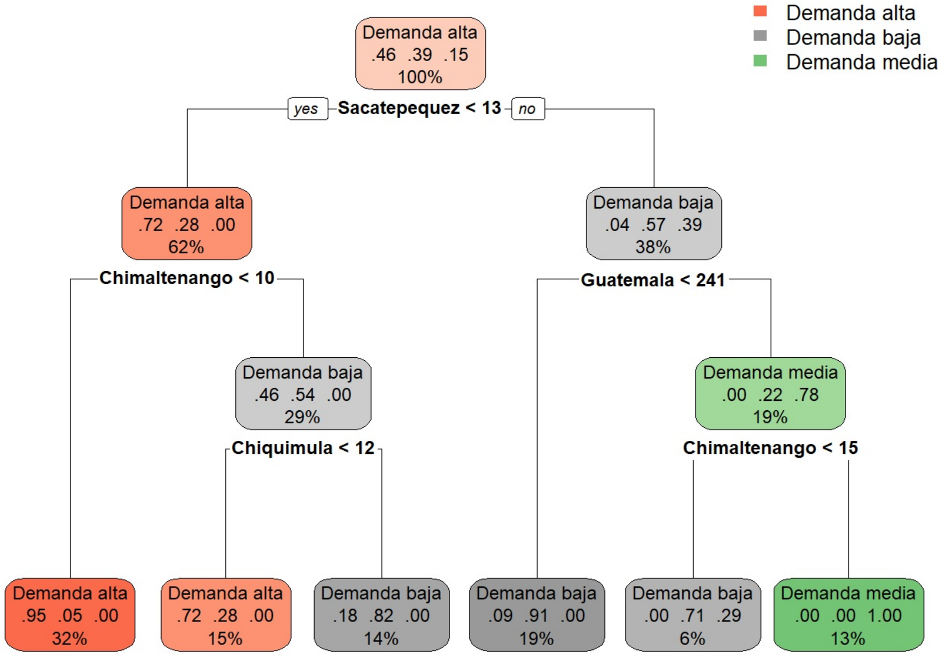
Grafica de Silueta (Divorcios por departamento por mes)

```
## [1] 0.7208329
```

Aplicacion de Naive Bayes (Divorcios por departamento por mes)

```
## + Fold01: usekernel= TRUE, fL=0, adjust=1
## - Fold01: usekernel= TRUE, fL=0, adjust=1
## + Fold01: usekernel=FALSE, fL=0, adjust=1
## - Fold01: usekernel=FALSE, fL=0, adjust=1
## + Fold02: usekernel= TRUE, fL=0, adjust=1
## - Fold02: usekernel= TRUE, fL=0, adjust=1
## + Fold02: usekernel=FALSE, fL=0, adjust=1
## - Fold02: usekernel=FALSE, fL=0, adjust=1
## + Fold03: usekernel= TRUE, fL=0, adjust=1
## - Fold03: usekernel= TRUE, fL=0, adjust=1
## + Fold03: usekernel=FALSE, fL=0, adjust=1
## - Fold03: usekernel=FALSE, fL=0, adjust=1
## + Fold04: usekernel= TRUE, fL=0, adjust=1
## - Fold04: usekernel= TRUE, fL=0, adjust=1
## + Fold04: usekernel=FALSE, fL=0, adjust=1
## - Fold04: usekernel=FALSE, fL=0, adjust=1
## + Fold05: usekernel= TRUE, fL=0, adjust=1
## - Fold05: usekernel= TRUE, fL=0, adjust=1
## + Fold05: usekernel=FALSE, fL=0, adjust=1
## - Fold05: usekernel=FALSE, fL=0, adjust=1
## + Fold06: usekernel= TRUE, fL=0, adjust=1
## - Fold06: usekernel= TRUE, fL=0, adjust=1
## + Fold06: usekernel=FALSE, fL=0, adjust=1
## - Fold06: usekernel=FALSE, fL=0, adjust=1
## + Fold07: usekernel= TRUE, fL=0, adjust=1
## - Fold07: usekernel= TRUE, fL=0, adjust=1
## + Fold07: usekernel=FALSE, fL=0, adjust=1
## - Fold07: usekernel=FALSE, fL=0, adjust=1
## + Fold08: usekernel= TRUE, fL=0, adjust=1
## - Fold08: usekernel= TRUE, fL=0, adjust=1
## + Fold08: usekernel=FALSE, fL=0, adjust=1
## - Fold08: usekernel=FALSE, fL=0, adjust=1
## + Fold09: usekernel= TRUE, fL=0, adjust=1
## - Fold09: usekernel= TRUE, fL=0, adjust=1
## + Fold09: usekernel=FALSE, fL=0, adjust=1
## - Fold09: usekernel=FALSE, fL=0, adjust=1
## + Fold10: usekernel= TRUE, fL=0, adjust=1
## - Fold10: usekernel= TRUE, fL=0, adjust=1
## + Fold10: usekernel=FALSE, fL=0, adjust=1
## - Fold10: usekernel=FALSE, fL=0, adjust=1
## Aggregating results
## Selecting tuning parameters
## Fitting fL = 0, usekernel = TRUE, adjust = 1 on full training set
## Confusion Matrix and Statistics
##
##                Reference
## Prediction      Demanda alta Demanda baja Demanda media
##   Demanda alta            14            2             0
##   Demanda baja             5            8             1
##   Demanda media            0            0             6
##
## Overall Statistics
##
##                Accuracy : 0.7778
##                  95% CI : (0.6085, 0.8988)
##     No Information Rate : 0.5278
##     P-Value [Acc > NIR] : 0.001786
##
##                   Kappa : 0.6444
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: Demanda alta Class: Demanda baja
## Sensitivity                       0.7368              0.8000
## Specificity                       0.8824              0.7692
## Pos Pred Value                    0.8750              0.5714
## Neg Pred Value                    0.7500              0.9091
## Prevalence                        0.5278              0.2778
## Detection Rate                    0.3889              0.2222
## Detection Prevalence              0.4444              0.3889
## Balanced Accuracy                 0.8096              0.7846
##                      Class: Demanda media
## Sensitivity                        0.8571
## Specificity                        1.0000
## Pos Pred Value                     1.0000
## Neg Pred Value                     0.9667
## Prevalence                         0.1944
## Detection Rate                     0.1667
## Detection Prevalence               0.1667
## Balanced Accuracy                  0.9286
```

Aplicacion de Arboles de Decision (Divorcios por departamento por mes)



Demanda alta
Demanda baja
Demanda media
Demanda alta .46 .39 .15 100%
yes
Sacatepequez < 13
no
Demanda alta .72 .28 .00 62%
Demanda baja .04 .57 .39 38%
Chimaltenango < 10
Guatemala < 241
Demanda baja .46 .54 .00 29%
Demanda media .00 .22 .78 19%
Chiquimula < 12
Chimaltenango < 15
Demanda alta .95 .05 .00 32%
Demanda alta .72 .28 .00 15%
Demanda baja .18 .82 .00 14%
Demanda baja .09 .91 .00 19%
Demanda baja .00 .71 .29 6%
Demanda media .00 .00 1.00 13%

```
##               7               9              11              16              18 
##  "Demanda alta"  "Demanda alta"  "Demanda baja"  "Demanda baja"  "Demanda alta" 
##              22              24              26              29              33 
##  "Demanda baja"  "Demanda alta"  "Demanda alta"  "Demanda baja"  "Demanda alta" 
##              35              38              43              47              48 
##  "Demanda alta"  "Demanda baja"  "Demanda alta"  "Demanda alta"  "Demanda alta" 
##              50              52              55              57              59 
##  "Demanda alta"  "Demanda alta"  "Demanda alta"  "Demanda baja"  "Demanda alta" 
##              67              68              69              70              72 
##  "Demanda alta"  "Demanda baja"  "Demanda baja"  "Demanda baja"  "Demanda alta" 
##              73              76              81              84              86 
##  "Demanda alta"  "Demanda baja"  "Demanda baja"  "Demanda alta"  "Demanda baja" 
##              92              93              96             107             110 
## "Demanda media"  "Demanda baja"  "Demanda baja"  "Demanda baja" "Demanda media" 
##             113 
## "Demanda media"
```

```
## Confusion Matrix and Statistics
## 
##                 Reference
## Prediction      Demanda alta Demanda baja Demanda media
##   Demanda alta            15            3             0
##   Demanda baja             0           14             1
##   Demanda media            0            0             3
## 
## Overall Statistics
##                                           
##                 Accuracy : 0.8889         
##                   95% CI : (0.7394, 0.9689)
##      No Information Rate : 0.4722         
##      P-Value [Acc > NIR] : 1.909e-07      
##                                           
##                    Kappa : 0.8103         
##                                           
##  Mcnemar's Test P-Value : NA             
## 
## Statistics by Class:
## 
##                       Class: Demanda alta Class: Demanda baja
## Sensitivity                        1.0000              0.8235
## Specificity                        0.8571              0.9474
## Pos Pred Value                     0.8333              0.9333
## Neg Pred Value                     1.0000              0.8571
## Prevalence                         0.4167              0.4722
## Detection Rate                     0.4167              0.3889
## Detection Prevalence               0.5000              0.4167
## Balanced Accuracy                  0.9286              0.8854
##                       Class: Demanda media
## Sensitivity                        0.75000
## Specificity                        1.00000
## Pos Pred Value                     1.00000
## Neg Pred Value                     0.96970
## Prevalence                         0.11111
## Detection Rate                     0.08333
## Detection Prevalence               0.08333
## Balanced Accuracy                  0.87500
## Confusion Matrix and Statistics
## 
##                 Reference
## Prediction      Demanda alta Demanda baja Demanda media
##   Demanda alta            15            6             0
##   Demanda baja             0           11             1
##   Demanda media            0            0             3
## 
## Overall Statistics
## 
##                 Accuracy : 0.8056
##                   95% CI : (0.6398, 0.9181)
##      No Information Rate : 0.4722
##      P-Value [Acc > NIR] : 4.227e-05
## 
##                    Kappa : 0.6706
## 
##  Mcnemar's Test P-Value : NA
## 
## Statistics by Class:
## 
##                       Class: Demanda alta Class: Demanda baja
## Sensitivity                        1.0000              0.6471
## Specificity                        0.7143              0.9474
## Pos Pred Value                     0.7143              0.9167
## Neg Pred Value                     1.0000              0.7500
## Prevalence                         0.4167              0.4722
## Detection Rate                     0.4167              0.3056
## Detection Prevalence               0.5833              0.3333
## Balanced Accuracy                  0.8571              0.7972
##                       Class: Demanda media
## Sensitivity                        0.75000
## Specificity                        1.00000
## Pos Pred Value                     1.00000
## Neg Pred Value                     0.96970
## Prevalence                         0.11111
## Detection Rate                     0.08333
## Detection Prevalence               0.08333
## Balanced Accuracy                  0.87500
```

Aplicacion de SVM (Divorcios por departamento por mes)

```r
library(caret)


set.seed(123)
porcentaje <- 0.7
corteDepar <- sample(nrow(divorciosDFdepar1), nrow(divorciosDFdepar1) * porcentaje)
trainDepar <- divorciosDFdepar1[corteDepar, ]
testDepar <- divorciosDFdepar1[-corteDepar, ]

modeloSVM <- train(Categoria ~ ., data = trainDepar, method = "svmLinear", trControl = trainControl(method = "cv"
, number = 10))

prediccionSVM <- predict(modeloSVM, newdata = testDepar[, 1:22])

# Generar la matriz de confusión
cmSVM <- confusionMatrix(prediccionSVM, testDepar$Categoria)

cmSVM
```

```
## Confusion Matrix and Statistics
## 
##                Reference
## Prediction      Demanda alta Demanda baja Demanda media
##   Demanda alta            14            0             0
##   Demanda baja             3           12             1
##   Demanda media            0            0             6
## 
## Overall Statistics
##                                           
##                Accuracy : 0.8889          
##                  95% CI : (0.7394, 0.9689)
##     No Information Rate : 0.4722          
##     P-Value [Acc > NIR] : 1.909e-07       
##                                           
##                   Kappa : 0.8252          
##                                           
##  Mcnemar's Test P-Value : NA              
## 
## Statistics by Class:
## 
##                      Class: Demanda alta Class: Demanda baja
## Sensitivity                       0.8235              1.0000
## Specificity                       1.0000              0.8333
## Pos Pred Value                    1.0000              0.7500
## Neg Pred Value                    0.8636              1.0000
## Prevalence                        0.4722              0.3333
## Detection Rate                    0.3889              0.3333
## Detection Prevalence              0.3889              0.4444
## Balanced Accuracy                 0.9118              0.9167
##                      Class: Demanda media
## Sensitivity                        0.8571
## Specificity                        1.0000
## Pos Pred Value                     1.0000
## Neg Pred Value                     0.9667
## Prevalence                         0.1944
## Detection Rate                     0.1667
## Detection Prevalence               0.1667
## Balanced Accuracy                  0.9286
```

Aplicacion de RANdomForest (Divorcios por departamento por mes)

```
library(caret)

set.seed(123)
porcentaje <- 0.7
corteDepar <- sample(nrow(divorciosDFdepar1), nrow(divorciosDFdepar1) * porcentaje)
trainDepar <- divorciosDFdepar1[corteDepar, ]
testDepar <- divorciosDFdepar1[-corteDepar, ]

modeloRF <- train(Categoria ~ ., data = trainDepar, method = "rf", trControl = trainControl(method = "cv", number
= 10))

prediccionRF <- predict(modeloRF, newdata = testDepar[, 1:22])

# Generar la matriz de confusión
cmRF <- confusionMatrix(prediccionRF, testDepar$Categoria)
cmRF
```

```
## Confusion Matrix and Statistics
## 
##                 Reference
## Prediction      Demanda alta Demanda baja Demanda media
##   Demanda alta            16            3             0
##   Demanda baja             1            9             1
##   Demanda media            0            0             6
## 
## Overall Statistics
##                                           
##                 Accuracy : 0.8611          
##                   95% CI : (0.705, 0.9533)
##      No Information Rate : 0.4722          
##      P-Value [Acc > NIR] : 1.413e-06       
##                                           
##                    Kappa : 0.7747          
##                                           
##  Mcnemar's Test P-Value : NA              
## 
## Statistics by Class:
## 
##                      Class: Demanda alta Class: Demanda baja
## Sensitivity                       0.9412              0.7500
## Specificity                       0.8421              0.9167
## Pos Pred Value                    0.8421              0.8182
## Neg Pred Value                    0.9412              0.8800
## Prevalence                        0.4722              0.3333
## Detection Rate                    0.4444              0.2500
## Detection Prevalence              0.5278              0.3056
## Balanced Accuracy                 0.8916              0.8333
##                      Class: Demanda media
## Sensitivity                        0.8571
## Specificity                        1.0000
## Pos Pred Value                     1.0000
## Neg Pred Value                     0.9667
## Prevalence                         0.1944
## Detection Rate                     0.1667
## Detection Prevalence               0.1667
## Balanced Accuracy                  0.9286
```