

Universidad de San Carlos de Guatemala
Escuela de estudios de postgrado
Maestría en ingeniería para la industria con especialización
en ciencias de la computación
Introducción a la minería de datos.
No carne: 999014035

Descripción

```
data$DEPARTAMENTO <- as.factor(data$DEPARTAMENTO)
```

```
set.seed(100)  
data <- data[sample(1:nrow(data)), ]
```

La variable objetivo, DEPARTAMENTO, se convierte en un factor, una codificación esencial para problemas de clasificación.

Los datos se reorganizan aleatoriamente (sample) para evitar sesgos en la partición.

```
index <- sample(1:nrow(data), 0.8 * nrow(data))  
train <- data[index, ]  
test <- data[-index, ]
```

Se dividen los datos en un 80% para entrenamiento y 20% para prueba, utilizando índices seleccionados aleatoriamente.

Entrenamiento del modelo:

Variable objetivo: DEPARTAMENTO (departamento asociado al migrante).

Predictoras:

PEI3: Sexo del migrante.

PEI4: Edad del migrante.

PEI5: Año de emigración.

Hiperparámetros:

ntree = 40: Número de árboles en el bosque.

mtry = 3: Número de variables consideradas al dividir un nodo.

```
bosque <- randomForest(DEPARTAMENTO ~ PEI3 + PEI4 + PEI5,  
                        data = train,  
                        ntree = 40,  
                        mtry = 3)
```

```

22
23 entreno <- predict(bosque, test)
24
25 dato_nuevo <- data.frame(PEI3=1, PEI4=50, PEI5=2004)
26
27 prediccion <- predict(bosque,dato_nuevo)
28 prediccion
29

```

21:33 (Top Level) ▾

Console Terminal x Background Jobs x

R 4.2.1 · ~/

```

> prediccion <- predict(bosque,dato_nuevo)
> prediccion
1
1
Levels: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
> dato_nuevo <- data.frame(PEI3=1, PEI4=50, PEI5=2004)
> prediccion <- predict(bosque,dato_nuevo)
> prediccion
1
13
Levels: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22

```

Se predice los datos para una persona con genero masculino, con 50 años de edad y que migro en 2004 y la predicción del modelo dio como resultado el número 13 de departamento que es el departamento de Huehuetenango según el diccionario de datos.

Etiqueta de los valores		
Nombre	Valor	Etiqueta
DEPARTAMENTO	1	Guatemala
	2	El Progreso
	3	Sacatepéquez
	4	Chimaltenango
	5	Escuintla
	6	Santa Rosa
	7	Sololá
	8	Totonicapán
	9	Quetzaltenango
	10	Suchitepéquez
	11	Retalhuleu
	12	San Marcos
	13	Huehuetenango
	14	Quiché
	15	Baja Verapaz
	16	Alta Verapaz
	17	Petén
	18	Izabal
	19	Zacapa
	20	Chiquimula
	21	Jalapa
	22	Jutiapa
	1	Urbano

Repositorio

<https://github.com/Mruiz-99/apriori>

En la carepta Tarea extra – comodín

Tarea por reemplazar

Tarea 3