

1. Variable predicha: sexo del agresor (AGR_SEXO)

En este modelo se utilizó un bosque aleatorio para clasificar el sexo del agresor (AGR_SEXO) a partir de la información disponible en el caso:

- Características de la víctima: edad (VIC_EDAD), escolaridad (VIC_ESCOLARIDAD), situación laboral (VIC_TRABAJA), ocupación (VIC_OCUP).
- Características del agresor: edad (AGR_EDAD), situación laboral (AGR_TRABAJA).
- Características del hecho: tipo de agresión (HEC_TIPAGRE).

AGR_SEXO se trató como una variable categórica con dos niveles según diccionario 1 Masculino, 2 Femenino.

El objetivo del modelo es entonces estimar la probabilidad de que el agresor pertenezca a cada categoría de sexo dado el perfil víctima agresor y el tipo de agresión.

Con la matriz de confusión y el cálculo de precisión:

- Precisión global aproximada del 80.2 % (pre = 0.8023), es decir, el modelo acierta el sexo del agresor en 8 de cada 10 casos de prueba.
- Es más preciso para la categoría 1 que para la 2, lo que indica que el modelo identifica mejor el patrón asociado al sexo 1 que al sexo 2.

2. Predicción concreta del caso “persona”

Para el caso hipotético:

```
persona <- data.frame(  
  VIC_EDAD = 30,  
  VIC_ESCOLARIDAD = 29,  
  VIC_TRABAJA = 2,  
  VIC_OCUP = 5153,  
  AGR_EDAD = 35,  
  AGR_TRABAJA = 1,  
  HEC_TIPAGRE = 3  
)  
result2 <- predict(bosque, persona, type = "prob")  
result2  
  1   2  
1 0.97 0.03
```

El modelo entrega:

- Probabilidad AGR_SEXO = 1: 0.97 (97 %)
- Probabilidad AGR_SEXO = 2: 0.03 (3 %)

Interpretación:

Dado el perfil de esta víctima (30 años, cierto nivel de escolaridad, no trabaja, ocupación 5153), el agresor (35 años, sí trabaja) y el tipo de agresión codificado como 3, el bosque aleatorio estima que hay un 97 % de probabilidad de que el agresor sea de sexo 1 y solo un 3 % de que sea de sexo 2.

Es decir, para este caso concreto el modelo está prácticamente seguro de que el agresor pertenece a la categoría 1 de AGR_SEXO.

Esta predicción se apoya en los patrones aprendidos del conjunto de datos 2024: en casos con características similares, la gran mayoría de agresores observados en los datos pertenece a esa misma categoría.

3. Gráfica del bosque aleatorio

El gráfico de plot(bosque) muestra el error OOB (out-of-bag) del modelo conforme aumenta el número de árboles:

- Al inicio, con pocos árboles, el error es más alto e inestable.
- A partir de 40–50 árboles, el error global se estabiliza cerca de 0.20, consistente con la precisión del 80 % calculada con el conjunto de prueba.
- La línea roja baja hasta 0.10 el modelo solo se equivoca alrededor del 10 % de las veces para la clase 1.
- La línea verde se mantiene alrededor de ~0.40 → el error para la clase 2 es bastante mayor (40 %), lo que confirma que la clase 2 es más difícil de predecir (o menos representada).

