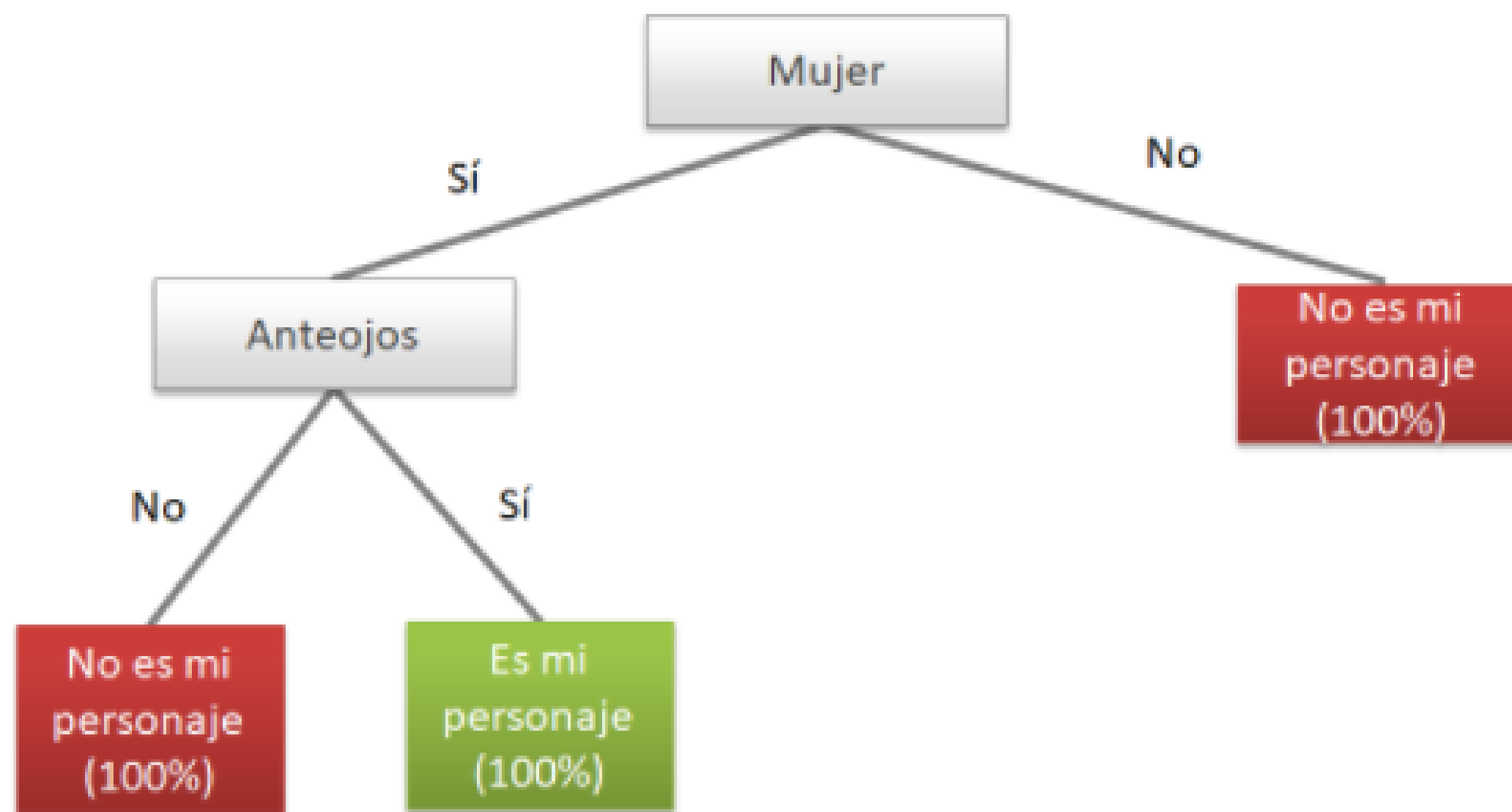


Tidymodels

Arboles de decision o Decision Trees

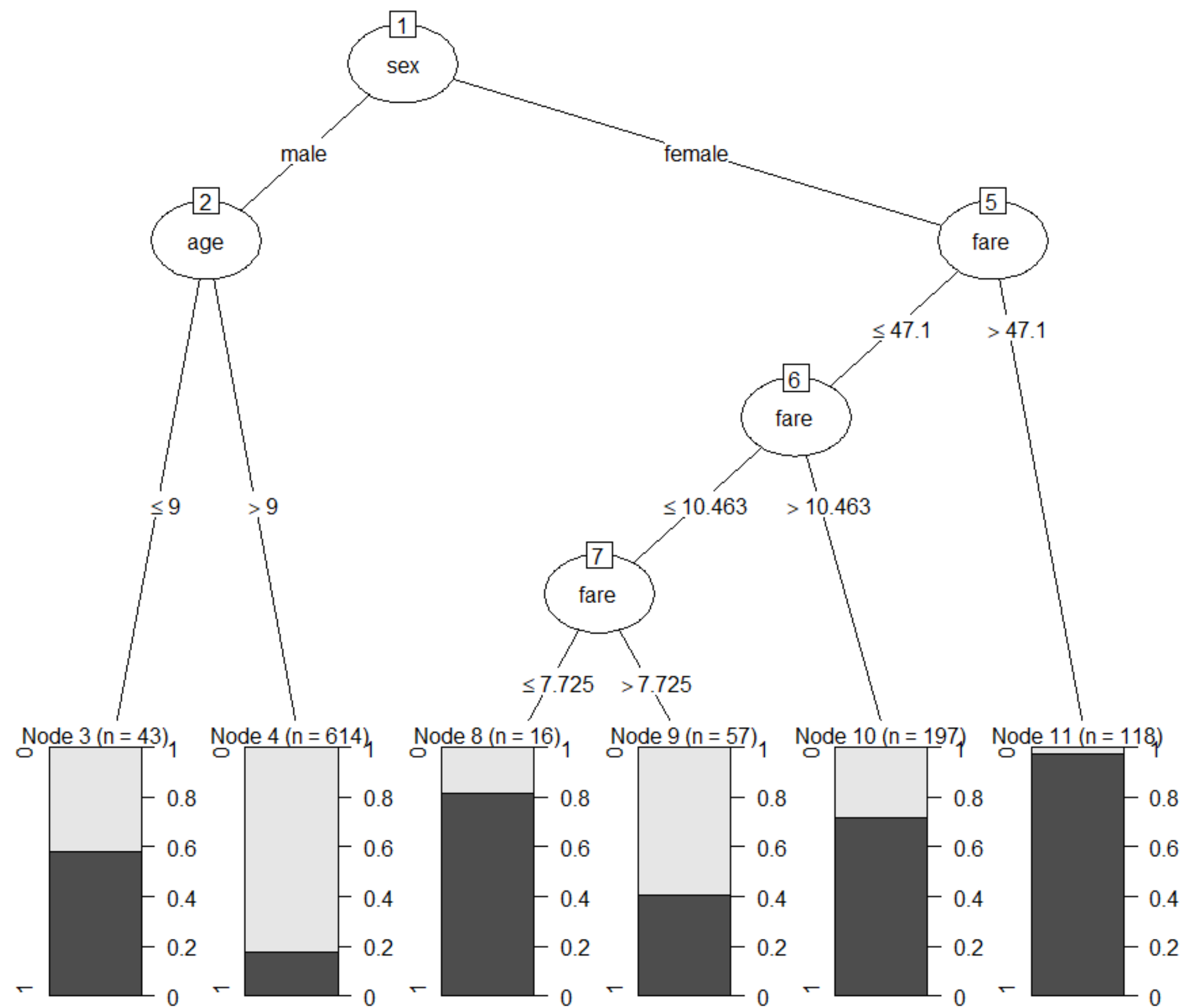
Modelo de clasificacion orientado a asignar una categoria a observaciones basados en variables independientes. Entonces, el modelo comienza con una variable dicotomica (si/no) y se va dividiendo en segmentos acorde a variables independientes que seleccionemos.

En este caso, al trabajar con un modelo de machine learning implica que no hay intervencion humana si no que se basa en las decisiones y reglas de la computadora para clasificar cada ves mejor.



A los datos!





Arboles de decision o Decision Trees

1. Nodos: el arbol esta compuesto de multiples nodos, los que estan al final cuando no hay mas segmentaciones se llaman hojas
2. Preguntas: hombre o mujer, valores de la tarifa, edad
3. Barras: muestran la proporcion que sobrevivio en gris oscuro y la que no sobrevivio (gris claro)

Arboles de decision o Decision Trees

Entropia: el arbol de decision selecciona cortes en nuestros datos que sean mas homogeneos en la clasificación. Es decir, se basa en que tan homogeneo es un conjunto de datos.

Mientras mas baja sea la entropia, mas homogeneidad hay en los datos.

En nuestro caso genero es la variable mas homogenea ya que tiene mayor cantidad de bifurcaciones.

Efectividad y Reglas

Luego de aplicar un modelo es necesario medir su efectividad. En el caso de un modelo de clasificación podemos utilizar una matriz de confusión.

Una de las principales ventajas de los árboles de decisión de este estilo es que podemos reducir su complejidad a un conjunto de reglas que nos permite clasificar los casos. Para esto solo tenemos que cambiar un parámetro al entrenar el árbol de decisión.

Esto es porque al no estar “obligado” a mostrar bifurcaciones en el árbol de decisión, lo que entrega son reglas y, al clasificar, elige la que tiene mayor accuracy

Gracias!

WWW.INSTRUMENTA.COM.AR | INFO@INSTRUMENTA.COM.AR | ☎ 351 800 20 40
AV. FIGUEROA ALCORTA 482 | CENTRO | X5000KFR | CÓRDOBA