Vamos a realizar un ejemplo completo de la construcción de un árbol de decisión utilizando la **reducción del error de clasificación** como criterio para dividir los datos en cada nodo.

Paso 1: Conjunto de Datos

Usaremos el siguiente conjunto de datos que representa si una persona comprará un producto basado en dos características: **Edad** (Joven, Adulto, Senior) y **Ingreso** (Alto, Bajo).

Edad	Ingreso	Compra
Joven	Bajo	No
Joven	Alto	Sí
Adulto	Bajo	No
Adulto	Alto	Sí
Senior	Bajo	No
Senior	Alto	No

En total, tenemos 6 instancias: 2 "Sí" y 4 "No".

Paso 2: Calcular el Error de Clasificación del Nodo Raíz

El **error de clasificación** se calcula como el porcentaje de ejemplos mal clasificados si todos fueran asignados a la clase mayoritaria del nodo. La fórmula es:

$$Error(S) = 1 - \max(p_{Sí}, p_{No})$$

En el nodo raíz:

- Hay 4 instancias de "No" y 2 instancias de "Sí".
- $p_{No} = \frac{4}{6} = 0.67, p_{Sí} = \frac{2}{6} = 0.33$

Por lo tanto, el error de clasificación del nodo raíz es:

$$Error(S) = 1 - 0.67 = 0.33$$

Paso 3: Calcular la Reducción del Error de Clasificación para Cada Atributo

Ahora calculamos el error de clasificación para cada atributo.

1. Dividir por el Atributo "Edad"

- **Edad = Joven**: 2 instancias (1 "Sí", 1 "No")
- **Edad = Adulto**: 2 instancias (1 "Sí", 1 "No")
- **Edad = Senior**: 2 instancias (0 "Sí", 2 "No")

Error de clasificación para cada subconjunto:

1. Para **Edad = Joven**: la mayoría de las instancias están equilibradas entre "Sí" y "No", así que la clasificación mayoritaria es cualquier clase (seleccionamos una arbitraria, como "No"):

$$Error(Joven) = 1 - \max(0.5, 0.5) = 1 - 0.5 = 0.5$$

2. Para **Edad = Adulto**: misma situación, instancias equilibradas entre "Sí" y "No":

$$Error(Adulto) = 1 - 0.5 = 0.5$$

3. Para **Edad = Senior**: la mayoría es "No" (2 "No"):

$$Error(Senior) = 1 - \max(0, 1) = 1 - 1 = 0$$

Error de clasificación ponderado para "Edad":

$$Error(Edad) = \frac{2}{6} \cdot 0.5 + \frac{2}{6} \cdot 0.5 + \frac{2}{6} \cdot 0 = 0.333$$

2. Dividir por el Atributo "Ingreso"

- **Ingreso = Bajo**: 3 instancias (0 "Sí", 3 "No")
- **Ingreso = Alto**: 3 instancias (2 "Sí", 1 "No")

Error de clasificación para cada subconjunto:

1. Para **Ingreso = Bajo**: todas las instancias son "No":

$$Error(Bajo) = 1 - \max(0, 1) = 1 - 1 = 0$$

2. Para **Ingreso = Alto**: la mayoría es "Sí" (2 "Sí", 1 "No"):

$$Error(Alto) = 1 - \max\left(\frac{2}{3}, \frac{1}{3}\right) = 1 - \frac{2}{3} = 0.333$$

Error de clasificación ponderado para "Ingreso":

$$Error(Ingreso) = \frac{3}{6} \cdot 0 + \frac{3}{6} \cdot 0.333 = 0.167$$

Paso 4: Comparar la Reducción del Error de Clasificación y Elegir el Mejor Atributo

Ahora comparamos los errores ponderados:

- **Edad** tiene un error ponderado de 0.333.
- **Ingreso** tiene un error ponderado de 0.167.

Como el error ponderado para **Ingreso** es menor, es el mejor atributo para dividir los datos.

Paso 5: Dividir los Datos por "Ingreso"

Al dividir por **Ingreso**, obtenemos dos subconjuntos:

1. **Ingreso = Bajo**: Todas las instancias son "No", por lo que este nodo es puro y no requiere más divisiones.
2. **Ingreso = Alto**: Este subconjunto tiene 2 "Sí" y 1 "No", por lo que aún se podría dividir por otro atributo (por ejemplo, **Edad**) si fuera necesario.

Resumen

El árbol de decisión elige el atributo **Ingreso** para la primera división porque minimiza el error de clasificación en los nodos hijos. A medida que avanzamos en la construcción del árbol, el algoritmo sigue seleccionando los atributos que proporcionan la mayor reducción del error de clasificación en cada paso.

Este proceso muestra cómo la **reducción del error de clasificación** se utiliza para tomar decisiones sobre cómo dividir los datos en un árbol de decisión, seleccionando los atributos que generan los nodos más puros.