

Árboles de decisión – Resumen

Los árboles de decisión son modelo de minería de datos predictivos que permite el análisis de datos a través de una estructura multivariable en forma de árbol. La figura 1 está la descripción de un árbol de decisión.

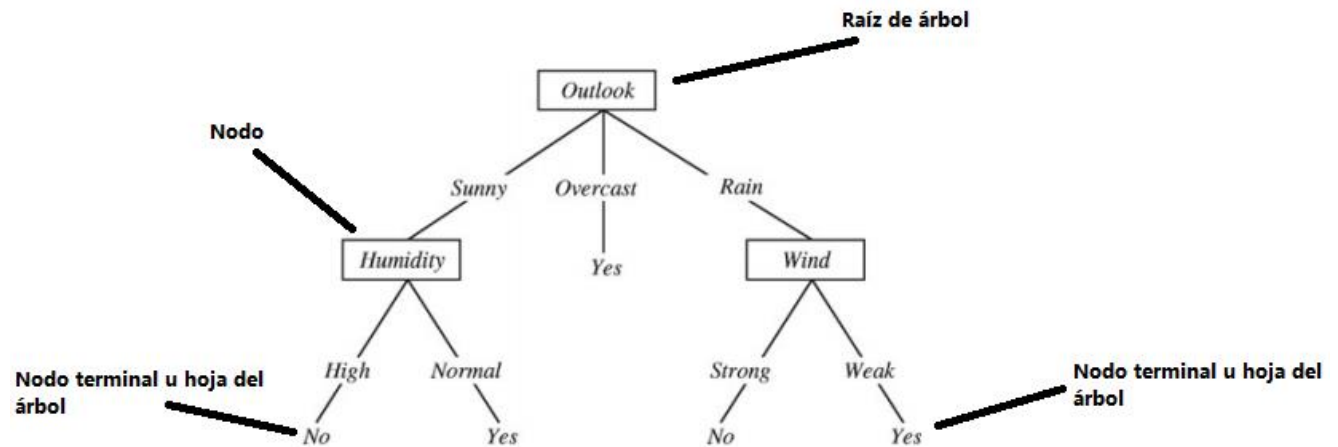


Figura 1 Estructura de un árbol de decisión

Existen dos tipos de árboles de decisión

- **Árbol de decisión *crisp*.**
 - Cuenta con un conjunto de datos de entrenamiento S .
 - Cada objeto en S es descrito por una lista de atributos A .
 - Cada elemento de A es representado por términos lingüísticos L .
 - La construcción de estos árboles de decisión involucra las siguientes decisiones:
 - Atributo por evaluar.
 - Condición de paro.
 - Asignación de nodos finales a cada clase.
 - El popular algoritmo *ID3* es uno de los más usados para árboles de decisión. Está descrito por la siguiente ecuación.

$$E(S) = \sum_{k=1}^K -p(k) \log_2 p(k) = \sum_{k=1}^K \frac{C_K}{|S|} \log_2 \frac{|C_K|}{|S|} \dots (1)$$

- Donde $E(S)$ es la entropía, $p(k)$ es la frecuencia relativa de la clase k es el conjunto S .

Alumno: Páez Ortega Oswaldo Emmanuel

Asignatura: Lógica difusa.

Grupo: 2TM11

- El pseudocódigo de el algoritmo *ID3* es el siguiente:

Entrada: Datos a clasificar

1 calcular $p(k)$, $E(S)$

2 determinar $IG(S, A_j) = \max_{i \in S} IG(S, A_i)$

3 ramificar cada atributo A_j

4 Actualizar S

5 si $E(S) == 0$ o ningún elemento queda por clasificar, termina la rama

6 si todas las ramas están terminadas, termina el algoritmo

7 si no, regresa paso 1

Salida: Árbol de decisión

- Otras medidas de ganancia de información son:
 - Impureza $GIN.I$
 - Error de clasificación errónea.
 - Índice de Goodman y Kruskan
 - Desviación
- Otros algoritmos de clasificación:
 - Algoritmo *C4.5*: Eficiente para clasificación de datos simbólicos.
 - Árboles de clasificación y regresión (*CART*): Hecho para árboles con números de hojas igual a 2.
 - Aprendizaje supervisado en búsqueda (*SLIQ*).

- **Árbol de decisión difuso.**

- A veces considerado como una generalización del caso básico.
- Algoritmos como *ID3* y *SLIQ* están disponibles en lógica difusa.
- Al igual que la lógica difusa, se usan funciones de membresía y sus traslapes para modelar algún fenómeno.
- Nuevamente, por el hecho de usar lógica difusa, se hacen uso de reglas de clasificación.