



# Instituto Tecnológico De Durango

---

Inteligencia Artificial

*Ingeniería en sistemas computacionales*

**NOMBRE DEL INTEGRANTE**

Oswaldo Zura Alvarado

Arboles de Decisiones

**Facilitador: Rodríguez Rivas José Gabriel**

**Grupo: 9AA**

**Victoria de Durango, Dgo.**

# Árboles de decisión

La inducción con árboles de decisión es uno de los métodos más sencillos y con más éxito para construir algoritmos de aprendizaje. Sirve como una buena introducción al área de aprendizaje inductivo, y es muy sencillo de implementar. Primero, se describen los árboles de decisión como herramienta de desarrollo, y luego se muestra cómo aprenderlos. Mientras tanto, se introducen ideas que aparecen en todas las áreas de aprendizaje inductivo.

## Árboles de decisión como herramienta de desarrollo

Un árbol de decisión toma como entrada un objeto o una situación descrita a través de un conjunto de atributos y devuelve una «decisión»: el valor previsto de la salida dada la entrada. Los atributos de entrada pueden ser discretos o continuos. A partir de ahora, asumiremos entradas discretas. El valor de la salida puede ser a su vez discreto o continuo; aprender una función de valores discretos se denomina clasificación; aprender una función continua se denomina regresión. Nos concentraremos en clasificaciones booleanas, en las cuales cada ejemplo se clasifica como verdadero (positivo) o falso (negativo).

Un árbol de decisión desarrolla una secuencia de test para poder alcanzar una decisión. Cada nodo interno del árbol corresponde con un test sobre el valor de una de las propiedades, y las ramas que salen del nodo están etiquetadas con los posibles valores de dicha propiedad. Cada nodo hoja del árbol representa el valor que ha de ser devuelto si dicho nodo hoja es alcanzado.

Un ejemplo sencillo se muestra en el problema de decidir si esperar por una mesa en un restaurante. Aquí, el propósito es aprender una definición para el predicado meta Esperar. Esto es un problema de aprendizaje, primero se tiene que establecer qué atributos están disponibles para describir los ejemplos del dominio. A partir de ahora, supondremos que vamos a decidir sobre la siguiente lista de atributos:

1. Alternativa: si existe un restaurante adecuado alternativo cerca.
2. Bar: si el restaurante tiene un área de bar confortable en la que se pueda esperar.
3. Vier/Sab: verdad si es viernes o sábado.
4. Hambriento: si estamos hambrientos.

5. Clientes: cuánta gente hay en el restaurante (puede tomar los valores: Ninguno, Algunos y Lleno).

6. Precio: el rango de precios del restaurante (\$, \$\$, \$\$\$).

7. Lloviendo: si está lloviendo fuera.

8. Reserva: si se ha hecho una reserva.

9. Tipo: la clase de restaurante (francés, italiano, tailandés, o hamburguesería).

10. TiempoEsperaEstimado: el tiempo de espera estimado por el dueño (0-10 minutos, 10-30 minutos, 30-60 minutos, >60).

En la Figura 18.2 se muestra el árbol de decisión que se ha utilizado. Téngase en cuenta que el árbol no usa los atributos Precio y Tipo, de hecho, se les considera irrelevantes.

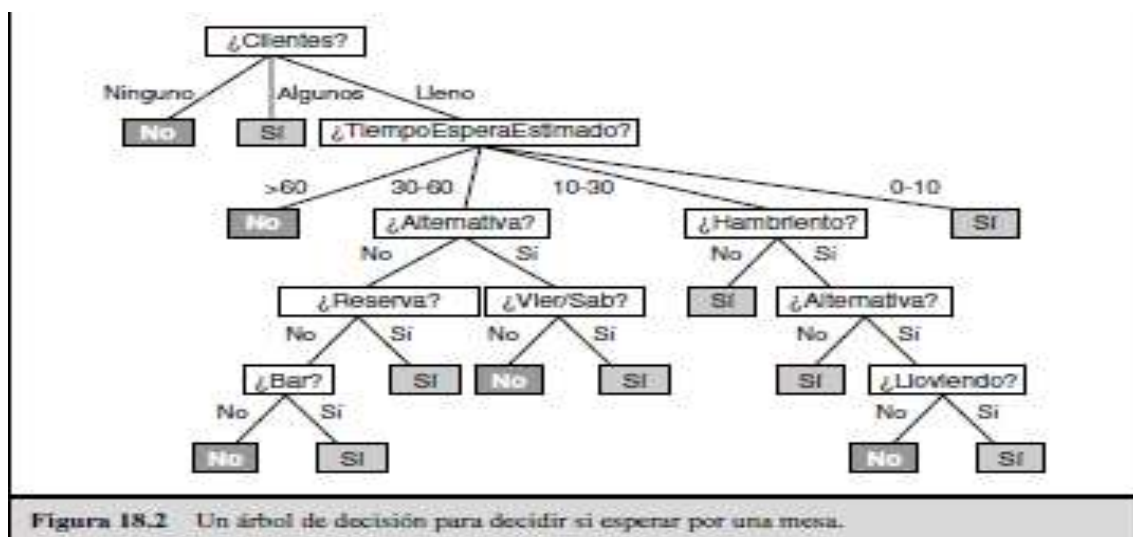


Figura 18.2 Un árbol de decisión para decidir si esperar por una mesa.

## Expresividad de los árboles de decisión

Desde el punto de vista de la lógica, cualquier hipótesis sobre el predicado meta Esperar, de un árbol de decisión particular, puede ser vista como una aserción de la forma:

$$\forall s \text{ Esperar}(s) \Leftrightarrow (P_1(s) \vee \dots \vee P_n(s))$$

donde cada condición  $P_i(s)$ , es una conjunción de test que corresponde a un camino desde la raíz del árbol hasta una hoja con un resultado positivo. Aunque ésta parezca una sentencia de primer orden, de alguna forma es proposicional, porque contiene exactamente una variable y todos los predicados son unarios.

## Ruido y sobreajuste

- **Datos perdidos:** Los valores pueden haberse perdido, o puede que sea muy caro obtenerlos. Esto da lugar a dos problemas: primero, dado un árbol de decisión completo, ¿cómo se debe clasificar un objeto para el que no se conoce uno de sus atributos de test? Segundo, ¿cómo se debe modificar la fórmula de la ganancia de información cuando algunos ejemplos tienen valores desconocidos para cierto atributo?
- **Atributos multivaluados:** cuando un atributo tiene muchos valores posibles, la ganancia de información proporciona una indicación inapropiada de su utilidad. En el caso extremo, se podría utilizar un atributo, como NombreRestaurante, que tenga un valor diferente para cada ejemplo.
- **Atributos de entrada continuos de valor entero:** los atributos continuos de valor entero, como Altura o Peso, tienen un número infinito de valores posibles. Más que generar infinitas ramas, los algoritmos de aprendizaje del árbol de decisión típicamente encuentran el punto de ruptura (Split point) que proporciona la máxima ganancia de información.
- **Atributos de salida de valor continuo:** Este tipo de árbol tiene en cada hoja una función lineal de algún subconjunto de atributos numéricos, en vez de un valor simple.

Un sistema de aprendizaje de árboles de decisión para aplicaciones reales debe ser capaz de manejar todos estos problemas. El manejo de variables de valor continuo es especialmente importante, porque tanto los procesos físicos como los financieros proporcionan datos numéricos.

En muchas áreas de industria y comercio, los árboles de decisión constituyen normalmente el primer método que se utiliza cuando el objetivo es extraer un método de clasificación del conjunto

de datos. Una propiedad importante de los árboles de decisión es que los humanos pueden entender la salida del algoritmo de aprendizaje.