

Árboles de decisión

Héctor Selley

Universidad Anáhuac México

20 de junio de 2023

1 ¿Qué es un árbol de decisión?

2 Ejemplos

- Ejemplo 1
- Ejemplo 2
- Ejemplo 3

¿Qué es un árbol de decisión?

¿Qué es un árbol de decisión?

- Un árbol de decisión es un modelo de predicción.

¿Qué es un árbol de decisión?

- Un árbol de decisión es un modelo de predicción.
- Se utiliza en diversas disciplinas como la Inteligencia Artificial, Medicina, Ingeniería, Ciencia de Datos y la Economía, entre muchas otras.

¿Qué es un árbol de decisión?

- Un árbol de decisión es un modelo de predicción.
- Se utiliza en diversas disciplinas como la Inteligencia Artificial, Medicina, Ingeniería, Ciencia de Datos y la Economía, entre muchas otras.
- Los árboles se construyen desde un conjunto de datos

¿Qué es un árbol de decisión?

- Un árbol de decisión es un modelo de predicción.
- Se utiliza en diversas disciplinas como la Inteligencia Artificial, Medicina, Ingeniería, Ciencia de Datos y la Economía, entre muchas otras.
- Los árboles se construyen desde un conjunto de datos
- Los diagramas resultantes son similares a los sistemas de predicción que se basan en reglas

¿Qué es un árbol de decisión?

- Un árbol de decisión es un modelo de predicción.
- Se utiliza en diversas disciplinas como la Inteligencia Artificial, Medicina, Ingeniería, Ciencia de Datos y la Economía, entre muchas otras.
- Los árboles se construyen desde un conjunto de datos
- Los diagramas resultantes son similares a los sistemas de predicción que se basan en reglas
- Sirven para categorizar una serie de condiciones que ocurren en forma sucesiva.

¿Qué es un árbol de decisión?

Los árboles de decisión se utilizan en cualquier proceso que implique una toma de decisión

¿Qué es un árbol de decisión?

Los árboles de decisión se utilizan en cualquier proceso que implique una toma de decisión, por ejemplo:

¿Qué es un árbol de decisión?

Los árboles de decisión se utilizan en cualquier proceso que implique una toma de decisión, por ejemplo:

- Búsqueda binaria

¿Qué es un árbol de decisión?

Los árboles de decisión se utilizan en cualquier proceso que implique una toma de decisión, por ejemplo:

- Búsqueda binaria
- Sistemas expertos

¿Qué es un árbol de decisión?

Los árboles de decisión se utilizan en cualquier proceso que implique una toma de decisión, por ejemplo:

- Búsqueda binaria
- Sistemas expertos
- Árboles de juego

¿Qué es un árbol de decisión?

- Los árboles de decisión son generalmente binarios

¿Qué es un árbol de decisión?

- Los árboles de decisión son generalmente binarios
- Significa que pueden tomar dos opciones

¿Qué es un árbol de decisión?

- Los árboles de decisión son generalmente binarios
- Significa que pueden tomar dos opciones
- Aunque es posible que existan árboles de tres o más opciones.

¿Para qué sirve un árbol de decisión?

Objetivos del árbol de decisión:

¿Para qué sirve un árbol de decisión?

Objetivos del árbol de decisión:

- Encontrar un árbol binario que clasifique datos de entrada con una *dispersión* mínima.

¿Para qué sirve un árbol de decisión?

Objetivos del árbol de decisión:

- Encontrar un árbol binario que clasifique datos de entrada con una *dispersión* mínima.
- Calcular la eficiencia del proceso de clasificación mediante la *dispersión*.

¿Qué es un árbol de decisión?

Árbol de decisión

Árbol de decisión es una técnica de estructura de datos jerárquicos que se utiliza para la clasificación y regresión de datos.

¿Qué es un árbol de decisión?

Árbol de decisión

Árbol de decisión es una técnica de estructura de datos jerárquicos que se utiliza para la clasificación y regresión de datos. Este método emplea la técnica *divide y vencerás*, mediante la cual encuentra recursivamente la separación por clasificación de los datos de entrada.

¿Qué es un árbol de decisión?

- Un árbol de decisión es un grafo que consiste en nodos y aristas.

¿Qué es un árbol de decisión?

- Un árbol de decisión es un grafo que consiste en nodos y aristas.
- Cada nodo puede tener máximo dos aristas, razón por lo que se le denomina como binario.

¿Qué es un árbol de decisión?

- Un árbol de decisión es un grafo que consiste en nodos y aristas.
- Cada nodo puede tener máximo dos aristas, razón por lo que se le denomina como binario.
- Un árbol de decisión responde una pregunta acerca de los datos y los clasifica de acuerdo con la respuesta de dicha pregunta.

¿Qué es un árbol de decisión?

- Un árbol de decisión es un grafo que consiste en nodos y aristas.
- Cada nodo puede tener máximo dos aristas, razón por lo que se le denomina como binario.
- Un árbol de decisión responde una pregunta acerca de los datos y los clasifica de acuerdo con la respuesta de dicha pregunta.

Utilizaremos algunos ejemplos para explicar los árboles de decisión, cómo se definen y construyen.

1 ¿Qué es un árbol de decisión?

2 Ejemplos

- Ejemplo 1
- Ejemplo 2
- Ejemplo 3

Ejemplo 1

La figura 1 muestra un árbol de decisión que mediante una pregunta, cuya respuesta puede ser verdadero o falso, clasifica los datos de entrada en dos grupos.

Ejemplo 1

La figura 1 muestra un árbol de decisión que mediante una pregunta, cuya respuesta puede ser verdadero o falso, clasifica los datos de entrada en dos grupos.

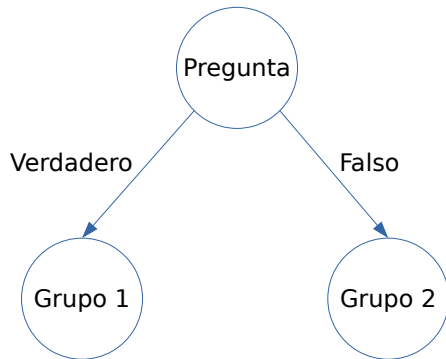


Figura: Ejemplo de árbol de decisión.

Ejemplo 1

- En los árboles, los **nodos** se representan con círculos o elipses en los cuales se aloja una pregunta

Ejemplo 1

- En los árboles, los **nodos** se representan con círculos o elipses en los cuales se aloja una pregunta
- Las aristas son la conexión entre ellos a través de la respuesta de la pregunta.
- Se denomina **rama** al conjunto de al menos dos nodos conectados por una arista.

Ejemplo 1

- Imagine que tiene un conjunto de datos que desea clasificar mediante una pregunta cuya respuesta es verdadero o falso. (figura 1)

Ejemplo 1

- Imagine que tiene un conjunto de datos que desea clasificar mediante una pregunta cuya respuesta es verdadero o falso. (figura 1)
- Esto permite clasificar los datos en dos grupos, uno cuya respuesta fue verdadera y otro cuya respuesta fue falsa.

Ejemplo 1

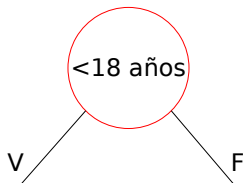
- Imagine que tiene un conjunto de datos que desea clasificar mediante una pregunta cuya respuesta es verdadero o falso. (figura 1)
- Esto permite clasificar los datos en dos grupos, uno cuya respuesta fue verdadera y otro cuya respuesta fue falsa.
- Para un árbol tan pequeño como el de este ejemplo, la separación de los datos es muy limitada por lo que se busca mejorarla empleando más nodos en el árbol, lo que significa un mayor número de categorías.

Ejemplo 1

- Imagine que tiene un conjunto de datos que desea clasificar mediante una pregunta cuya respuesta es verdadero o falso. (figura 1)
- Esto permite clasificar los datos en dos grupos, uno cuya respuesta fue verdadera y otro cuya respuesta fue falsa.
- Para un árbol tan pequeño como el de este ejemplo, la separación de los datos es muy limitada por lo que se busca mejorarla empleando más nodos en el árbol, lo que significa un mayor número de categorías.
- Adicionalmente, la pregunta sólo acepta respuestas absolutas, si se requiere de un rango de respuestas, por ejemplo, un rango de números habría que modificar el árbol.

Ejemplo 2

Construyamos un árbol con más nodos y ramificaciones, para este ejemplo se clasifica una persona de acuerdo con su edad. Se clasifica a una persona como adulto si su edad es mayor o igual a 18 años, como adolescente si está entre los 12 y 18 años, como niño si está entre los 2 y 12 años y como bebé si es menor a 2 años. El árbol resultante se muestra en la figura 2. En el árbol resultante de la figura 2 clasifica a las personas de acuerdo con su edad utilizando el criterio antes mencionado. En el árbol las personas han sido clasificadas en los nodos: adulto, adolescente, niño y bebé. A estos nodos, aquellos que tienen flechas que llegan a él pero no salen de él, se les denomina como **nodos terminales** o de decisión. Al nodo inicial, aquel del cual sólo salen flechas de él pero no entran, se le denomina **nodo raíz** o simplemente **raíz**. Los demás simplemente se les denomina como **nodos**.



Ejemplo 2

En la figura 2 el nodo en rojo es la raíz, los nodos en verde son terminales y los azules son simplemente nodos.

En el árbol de decisión de la figura 2 clasifica adecuadamente a las personas, dado que una persona sólo tiene una edad, la clasificación es perfecta de esa forma. Imagínese que deseamos clasificar personas de acuerdo con otro criterio, un criterio en el cual la respuesta no será tan específica como la edad o incluso puede que no haya una respuesta. Por ejemplo, imagine que deseamos clasificar personas de acuerdo con su sabor preferido de helado, puede que tenga uno, varios o incluso ninguno. En una situación como ésta, habrá una **impureza** en la clasificación. Más adelante se explicará a través del ejemplo la impureza y cómo se calcula.

Ejemplo 3

