

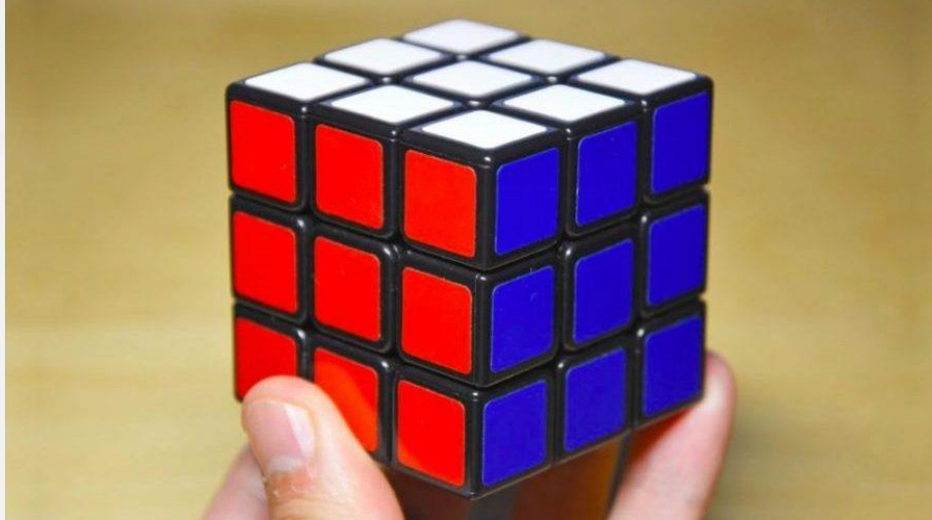
Arboles de Decisión



Ing. MBA. Julián Andrés Quimbayo
Docente Tiempo Completo
Programa Ingeniería de Sistemas



Divide y Vencerás



Arboles de Decisión

En general, los árboles de decisión aplican una estrategia similar de dividir los datos en porciones cada vez más pequeñas para identificar patrones que puedan utilizarse para la predicción. El conocimiento se presenta entonces en forma de estructuras lógicas que pueden entenderse sin ningún conocimiento estadístico. Especialmente útiles para la estrategia empresarial y la mejora de procesos



Tipos de Árboles de Decisión

Los tipos de árbol de decisión se basan en el tipo de variable objetivo que tenemos. Puede ser de dos tipos:

Árbol de decisión de variable categórica: El árbol de decisión que tiene una variable objetivo categórica se llama árbol de decisión de variable categórica. Ejemplo: En el escenario anterior del problema del estudiante, donde la variable objetivo era "El estudiante jugará al cricket o no", es decir, SÍ o NO.

Árbol de decisión de variable continua: El árbol de decisión tiene una variable objetivo continua, entonces se llama árbol de decisión de variable continua.

Terminología

Veamos la terminología básica utilizada con los árboles de decisión:

Nodo raíz: Representa toda la población o muestra y ésta se divide a su vez en dos o más conjuntos homogéneos.

División: Es un proceso de división de un nodo en dos o más subnodos.

Nodo de decisión: Cuando un subnodo se divide en otros subnodos, se denomina nodo de decisión.

Nodo de hoja/terminal: Los nodos que no se dividen se denominan nodo hoja o nodo terminal.

Terminología

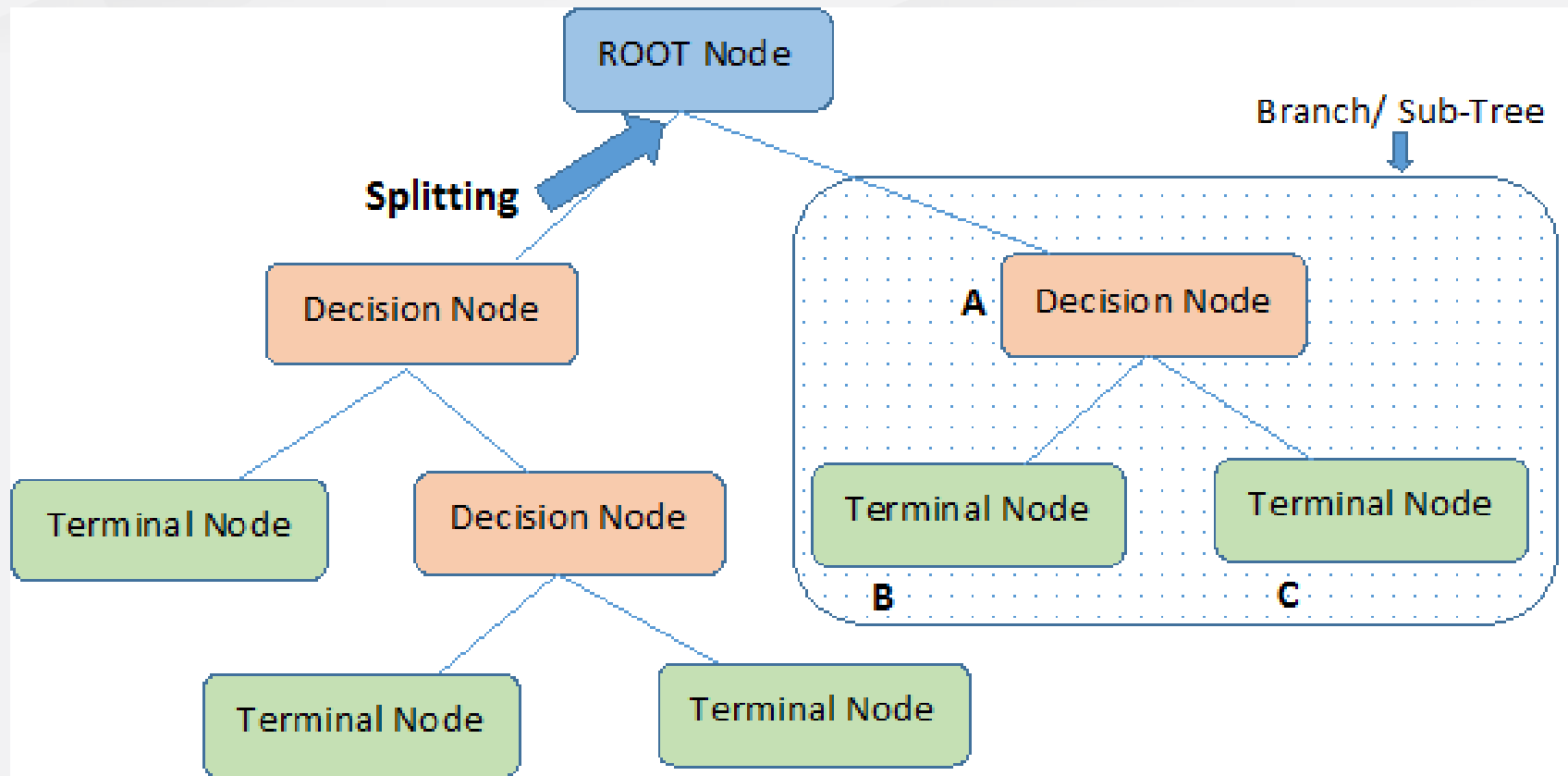
Veamos la terminología básica utilizada con los árboles de decisión:

poda: Cuando eliminamos los subnodos de un nodo de decisión, este proceso se llama poda. Se puede decir que es el proceso opuesto a la división.

Rama / Subárbol: Una sub-sección de todo el árbol se llama rama o sub-árbol.

Nodo padre e hijo: Un nodo dividido en subnodos se denomina nodo padre de subnodos y los subnodos son hijos del nodo padre. Estos son los términos comúnmente utilizados para los árboles de decisión. Como sabemos que cada algoritmo tiene ventajas y desventajas, a continuación se indican los factores importantes que se deben conocer.

Terminología

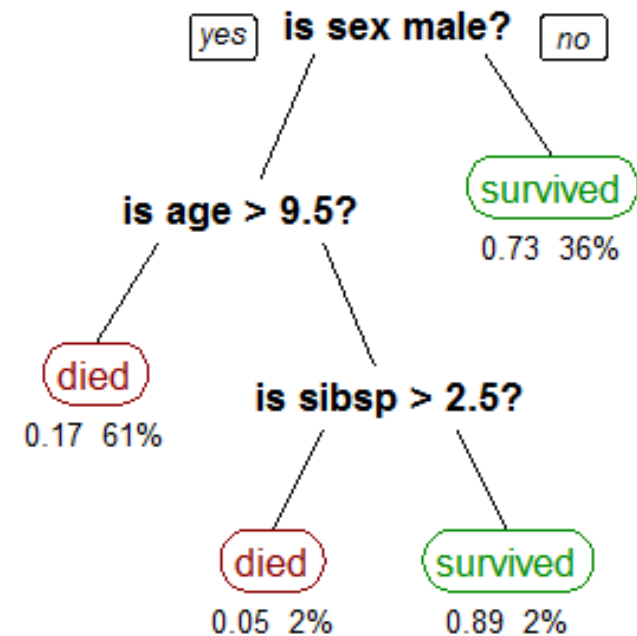


Note:- A is parent node of B and C.

Entendiendo los Árboles de Decisión

El modelo en sí comprende una serie de decisiones lógicas, similares a un diagrama de flujo, con nodos de decisión que indican una decisión a tomar sobre un atributo. Estos se dividen en ramas que indican las opciones de la decisión. El árbol termina con nodos de hoja (también conocidos como nodos terminales) que denotan el resultado de seguir una combinación de decisiones.

Los datos que se van a clasificar comienzan en el nodo raíz, donde pasan por las decisiones del árbol en función de los valores de sus características. El camino que recorren los datos lleva a cada registro a un nodo hoja, que le asigna una clase prevista.

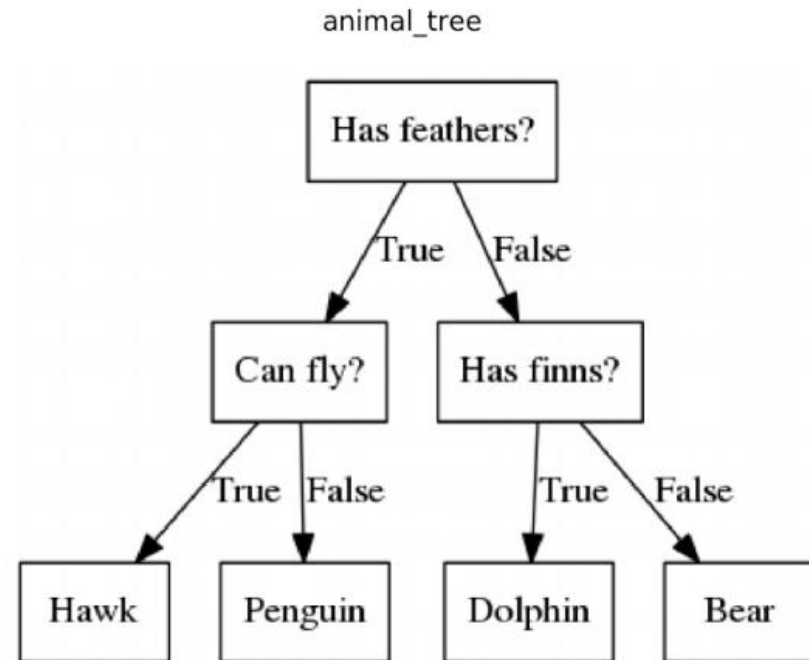


Entendiendo los Árboles de Decisión

Modelos de puntuación de crédito en los que los criterios que hacen que un solicitante sea rechazado deben estar bien especificados.

Estudios de marketing sobre la pérdida de clientes o su satisfacción que sean compartidos con la dirección o las agencias de publicidad.

Diagnóstico de enfermedades basado en mediciones de laboratorio, los síntomas o el índice de progresión de la enfermedad.



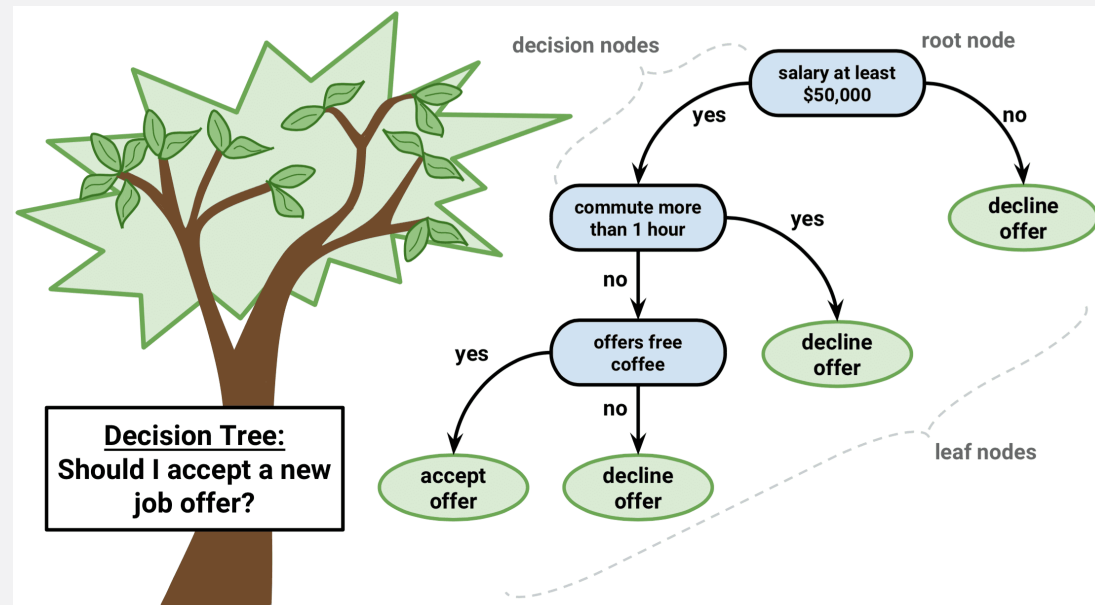
Entendiendo los Árboles de Decisión

Los árboles de decisión se construyen utilizando una heurística llamada partición recursiva. Este enfoque se conoce generalmente como divide y vencerás porque utiliza los valores de las características para dividirlos datos en subconjuntos cada vez más pequeños de clases similares. Empezando por el nodo raíz, que representa todo el conjunto de datos, el algoritmo elige una característica que sea la más predictiva de la clase objetivo.

Todos (o casi todos) los ejemplos del nodo tienen la misma clase.

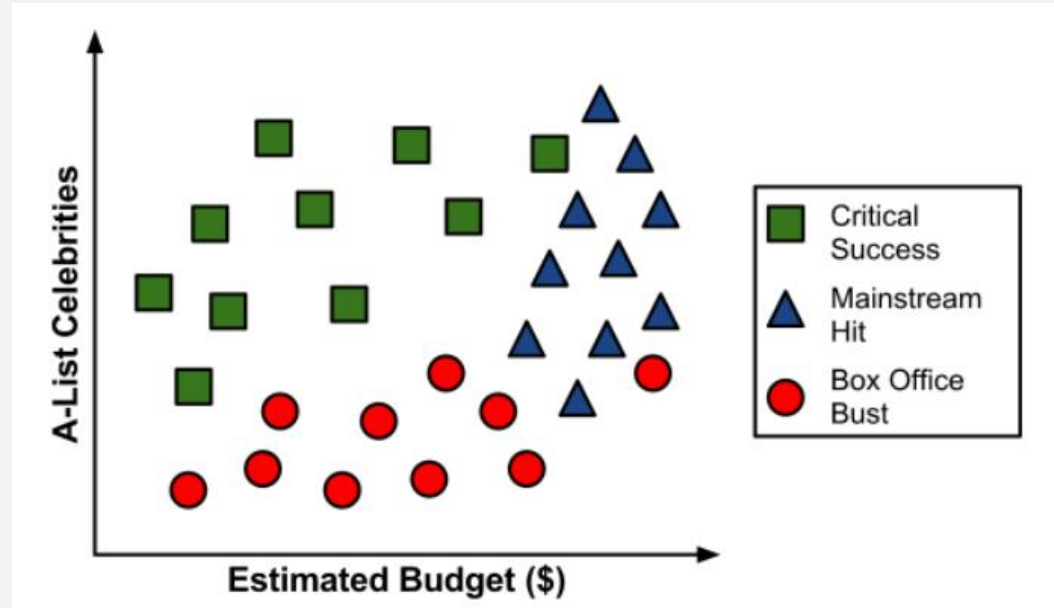
No hay características restantes para distinguir entre los ejemplos.

El árbol ha crecido hasta un límite de tamaño predefinido.

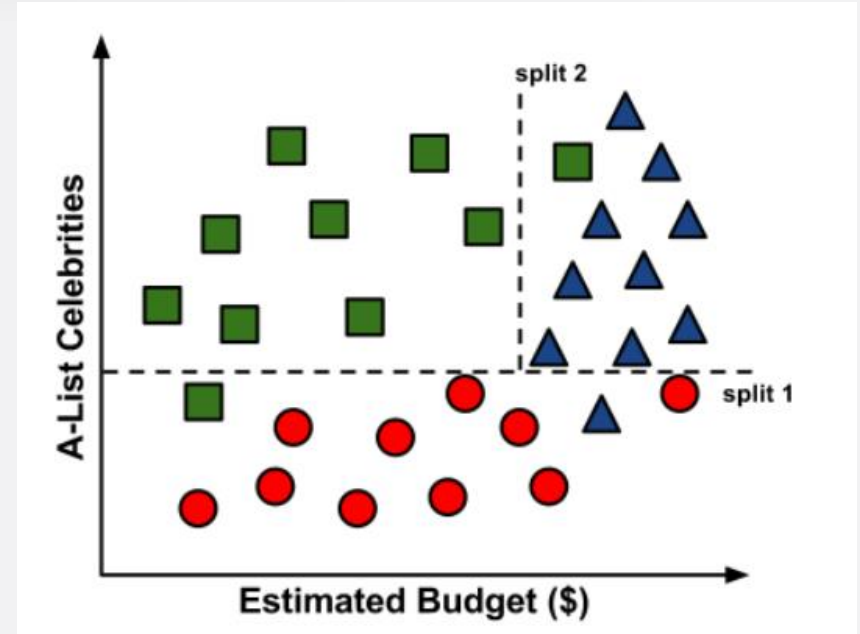
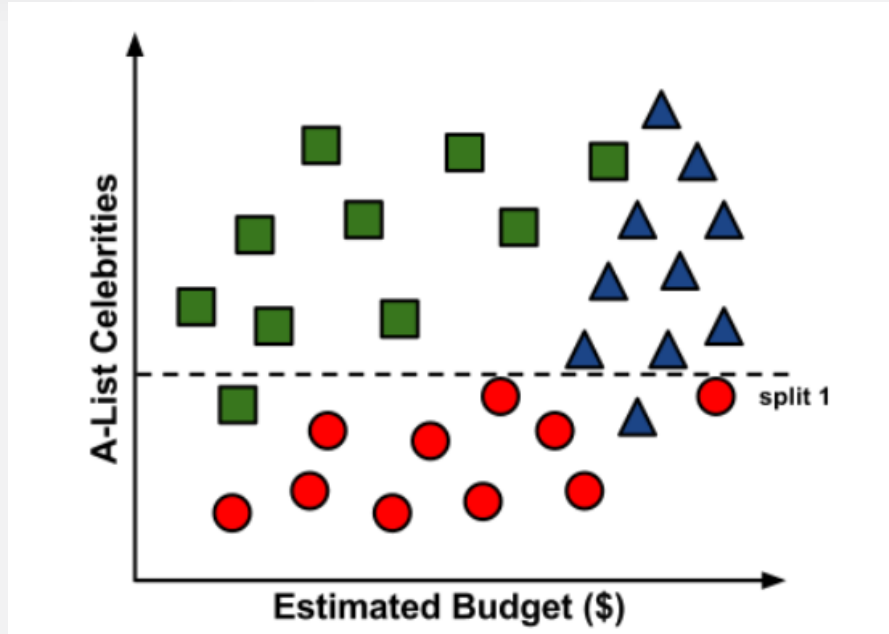


Entendiendo los Árboles de Decisión

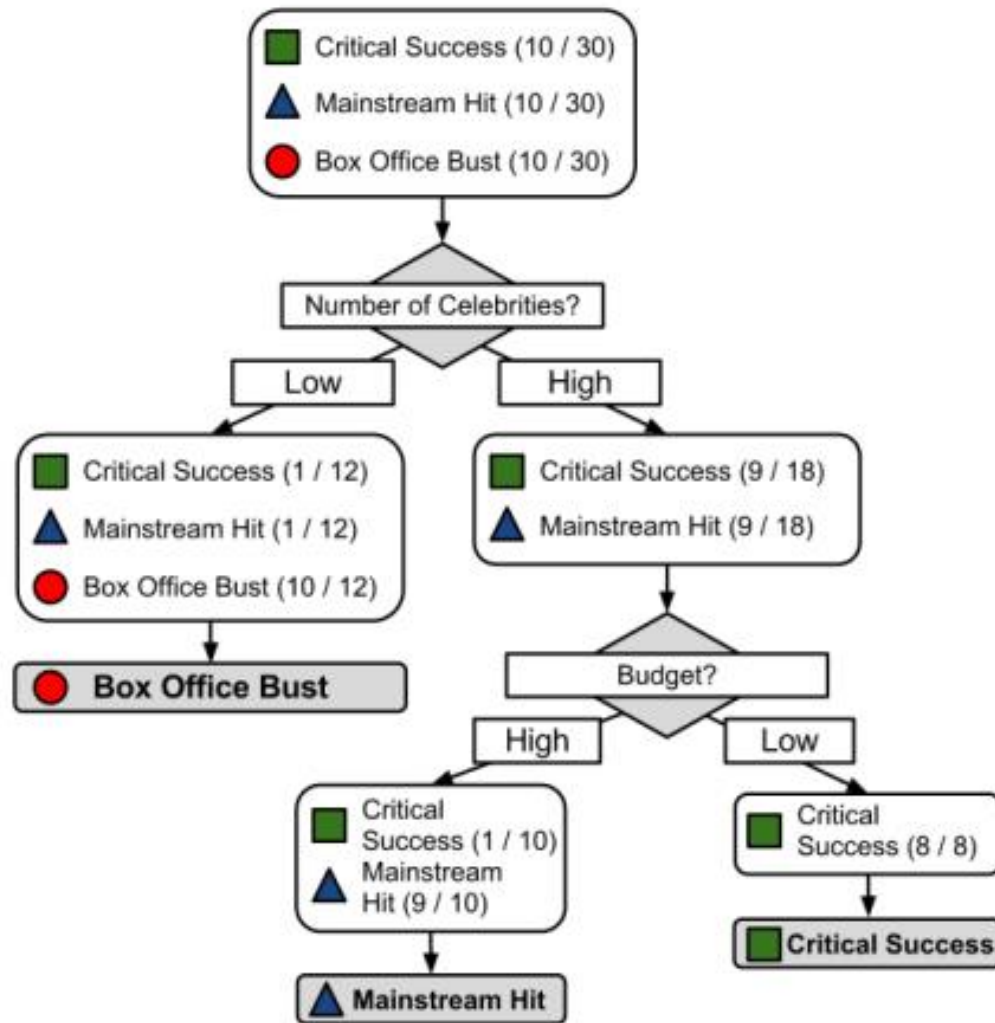
Para ilustrar el proceso de construcción del árbol, consideremos un ejemplo sencillo. Imagínesse que trabaja en un estudio cinematográfico de Hollywood y su mesa está repleta de guiones. En lugar de leer cada uno de ellos de principio a fin, decide desarrollar un algoritmo de árbol de decisión para predecir si una película potencial entraría en una de tres categorías: éxito de ventas, elección de la crítica o fracaso de taquilla. Para recopilar datos para tu modelo, te diriges a los archivos de los estudios para examinar los diez años anteriores de estrenos de películas.



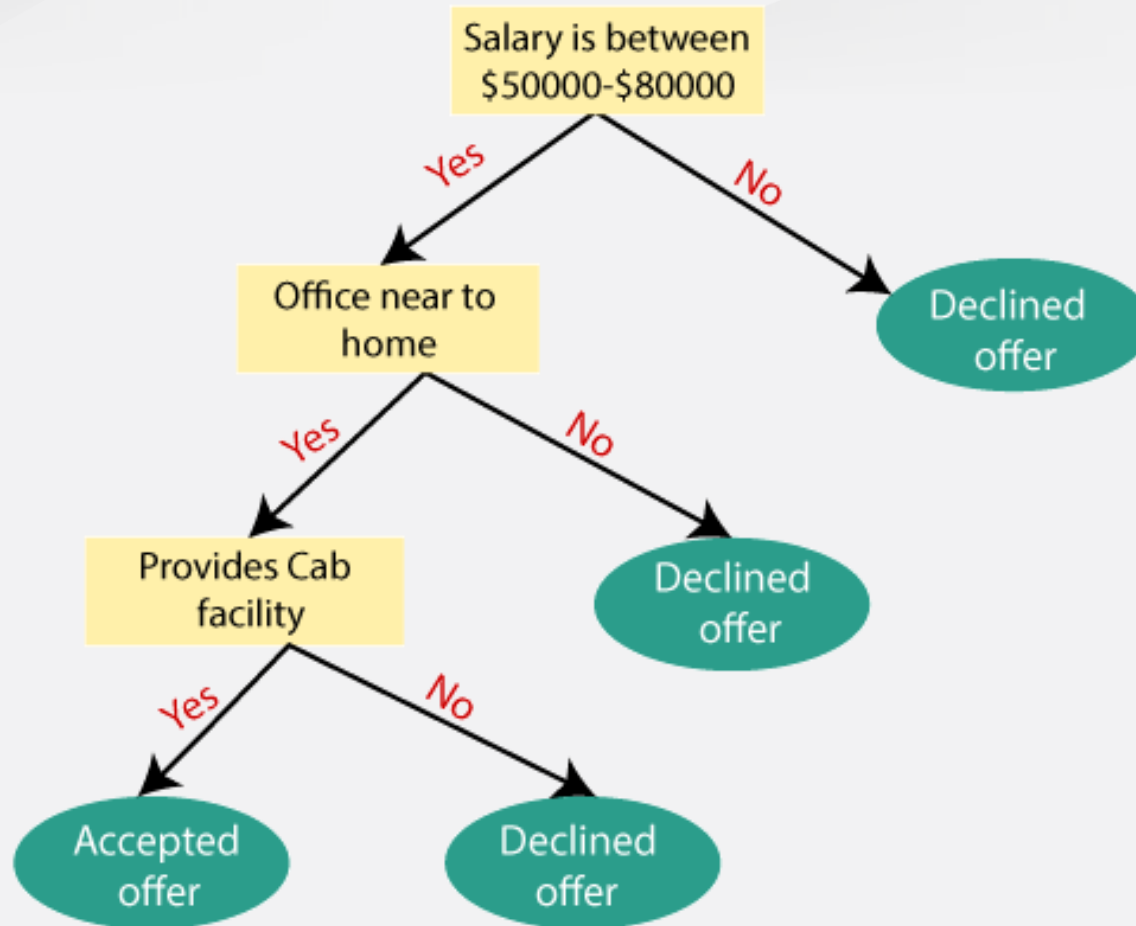
Entendiendo los Árboles de Decisión



Entendiendo los Árboles de Decisión



Entendiendo los Árboles de Decisión



¿Regresión o Árboles de Decisión?

En realidad, se puede utilizar cualquier algoritmo. Depende del tipo de problema que esté resolviendo. Veamos algunos factores clave que le ayudarán a decidir qué algoritmo utilizar:

Si la relación entre la variable dependiente e independiente está bien aproximada por un modelo lineal, la regresión lineal superará al modelo basado en árboles.

Si la relación entre las variables dependientes e independientes es muy compleja y no lineal, el modelo de árbol superará al método de regresión clásico.

Si necesita construir un modelo que sea fácil de explicar a la gente, un modelo de árbol de decisión siempre será mejor que un modelo lineal. Los modelos de árboles de decisión son incluso más sencillos de interpretar que la regresión lineal.

Bibliografía

Advanced R Statistical Programming and Data Models: Analysis, Machine Learning, and Visualization Matt Wiley, Joshua F. Wiley

<https://b-ok.lat/book/3698554/995fc6>

Machine Learning Using R: With Time Series and Industry-Based Use Cases in R Karthik Ramasubramanian, Abhishek Singh

<https://b-ok.lat/book/3642842/4ad30e>

Data Science with R A Step By Step Guide With Visual Illustrations and Examples Andrew Oleksy

<https://b-ok.lat/book/3640664/bcba3a>

Fuente: <https://es.z-lib.org/>


MUCHAS GRACIAS!


Julián Andrés Quimbayo Castro
MBA. En Dirección de Proyectos
Ingeniero de Sistemas



Más de 8 años de experiencia en el sector educativo y empresarial enseñando temas relacionados con el desarrollo de software (Java, C++, Node.js y PHP), proyectos y ciencia de datos con R y Python. Docente tiempo completo Corporación Universitaria del Huila – Corhuila.

Si deseas trabajar en equipo:

 julian.quimbayo@corhuila.edu.co

 Perfil linkedIn

“Si perdemos todo el presente criticando el pasado, nos olvidamos de estar presentes para crear el futuro.”

Alex Rovira Celma

