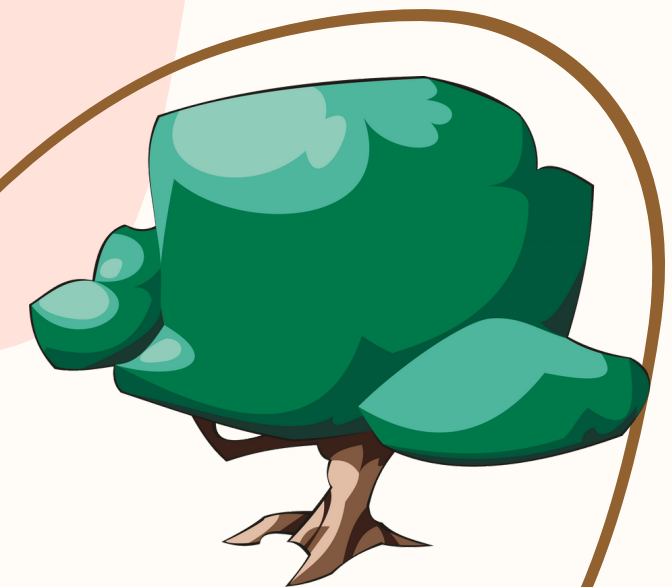
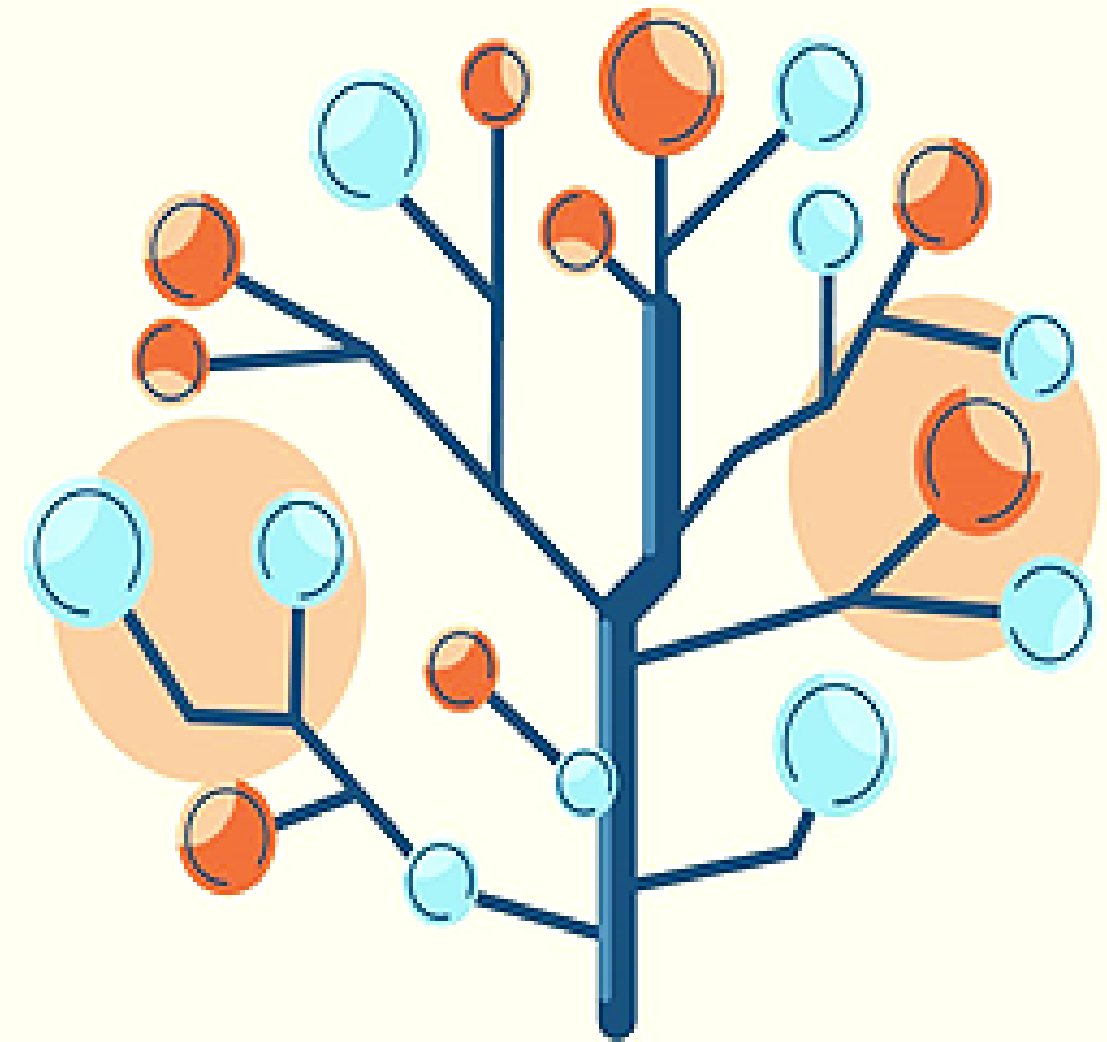


Árboles de decisión



¿Qué son los árboles de decisión?

Es un modelo de predicción utilizado en el ámbito de la inteligencia artificial, en términos técnicos es un diagrama que representa en forma secuencial condiciones y acciones; muestra qué condiciones se consideran en primer lugar, en segundo lugar y así sucesivamente para tomar una "acción"



ANTECEDENTES

01

El uso de árboles de decisión tuvo su origen en los trabajos de Sonquist y Morgan en el año 1964 y Morgan y Messenger en el año 1979, ambos realizados en la Universidad de Michigan.

02

El programa para la "Detección de Interacciones Automáticas", creada el año 1971 por los investigadores Sonquist, Baker y Morgan, fue uno de los primeros métodos de ajuste de los datos basados en árboles de clasificación.

03

Hacia el año 1984, los investigadores Breiman, Friedman, Olshen y Stone, introdujeron un nuevo algoritmo para la construcción de árboles y los aplicaron a problemas de regresión y clasificación.

04

El método es conocido como "Árboles de clasificación y regresión". Casi al mismo tiempo el proceso de inducción mediante árboles de decisión comenzó a ser usado por más gente.

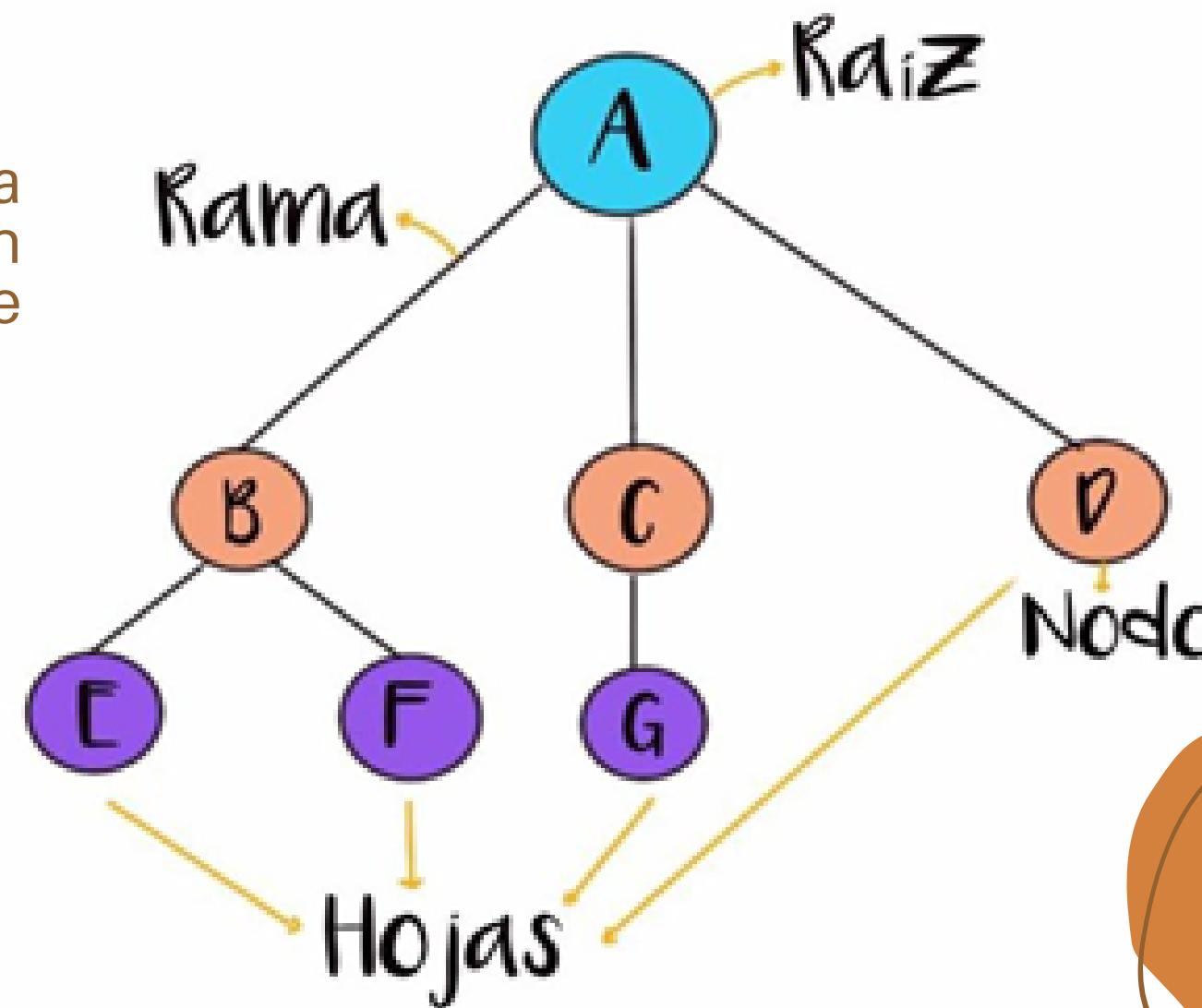
ESTRUCTURA DE LOS ÁRBOLES DE DECISION

nodos
raiz

En él se produce la primera división en función de la variable más importante.

nodos
intermedios

Tras la primera división encontramos estos nodos, que vuelven a dividir el conjunto de datos en función de las variables.



Determina el número máximo de nodos de una rama.

profundidad
del árbol

ramas

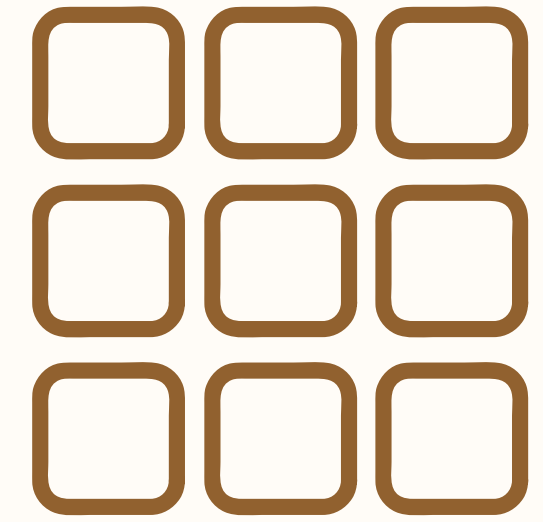
Brindan los posibles caminos que se tienen de acuerdo con la decisión tomada

hojas

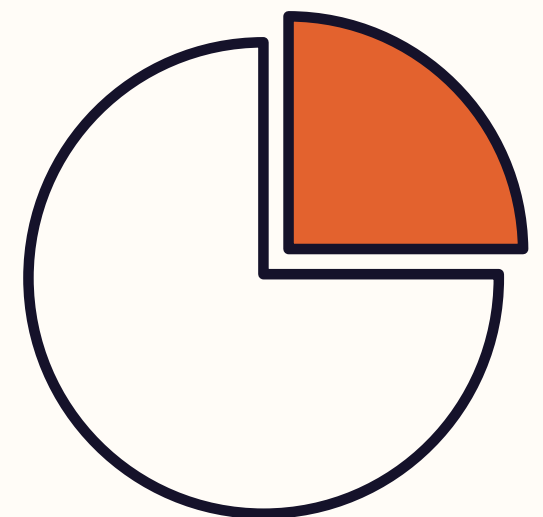
Se ubican en la parte inferior del esquema y su función es indicar la clasificación definitiva



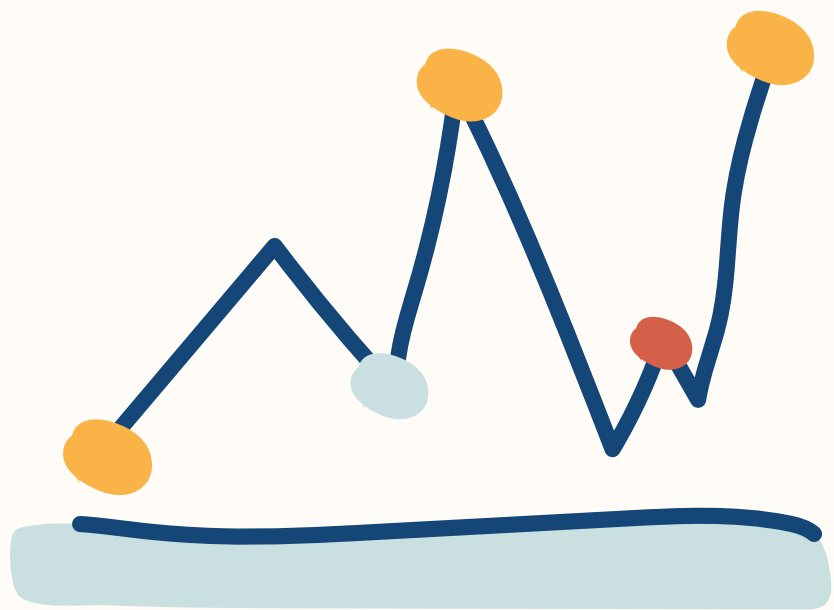
VENTAJAS



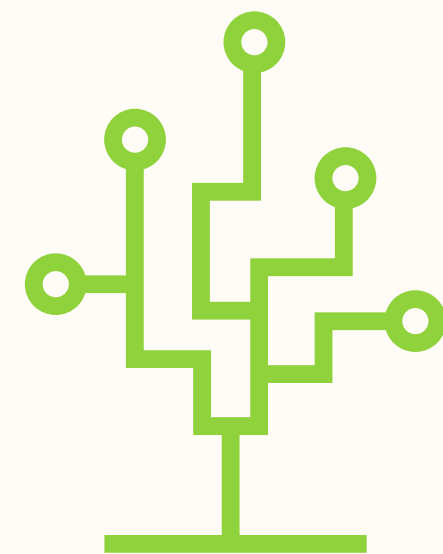
- Son fáciles de construir, interpretar y visualizar.
- Selecciona las variables más importantes y en su creación no siempre se hace uso de todos los predictores.
- Nos podemos servir de ellos para categorizar variables numéricas.



DESVENTAJAS



- Tienden al sobreajuste u overfitting de los datos, por lo que el modelo al predecir nuevos casos no estima con el mismo índice de acierto.
- Se ven influenciadas por los outliers, creando árboles con ramas muy profundas que no predicen bien para nuevos casos. Se deben eliminar dichos outliers.
- No suelen ser muy eficientes con modelos de regresión.



EJEMPLO



AaronAlvz08/MineriaDatos003

Contribute to AaronAlvz08/MineriaDatos003 development by creating an account on GitHub.



GitHub

BIBLIOGRAFIA

https://www.cienciadedatos.net/documentos/py07_arboles_decision_python.html

https://scikit-learn.org/stable/datasets/toy_dataset.html

Descarga de librerías desde el "cmd"

conda install scikit-learn

conda install -c anaconda python-graphviz

pip install matplotlib



Gracias por su atencion

EQUIPO 7

Paola Ibarra Medrano 1731502

Reyna Yarelli Blanco Gonzalez 1744128

Marissa Aleida Lucas Belmar 1799361

Aaron Emmanuel Garza Alvarez 1723466