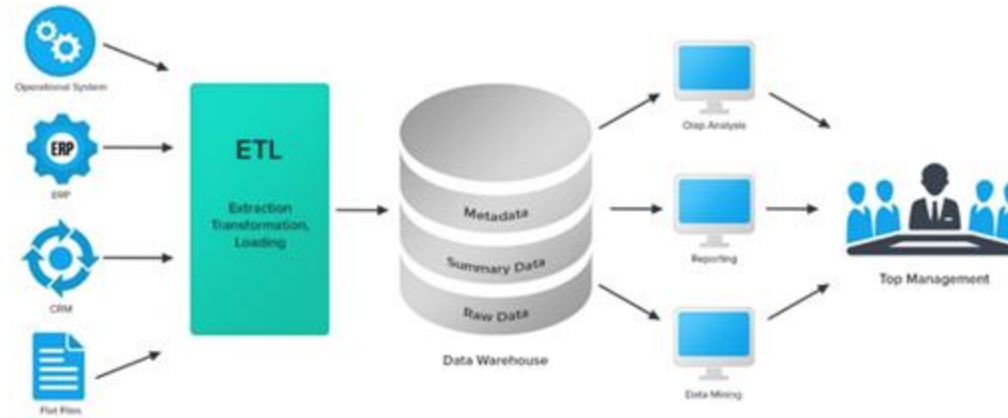


FACULTAD DE INGENIERÍA ESCUELA DE COMPUTACIÓN



Datawarehouse y Minería de Datos

Guía #11: Árbol de decisión

Árbol de decisión

Los árboles de decisión son un tipo de algoritmo que clasifica la información de forma que, como resultado, se genere un modelo en forma de árbol. Se trata de un modelo esquematizado de la información que representa las diferentes alternativas junto con los posibles resultados para cada alternativa elegida.

Los árboles de decisión son un tipo de modelo muy utilizado debido a que facilita mucho la comprensión de las diferentes opciones.

Por ejemplo, son muy utilizados en sectores como la banca y las aseguradoras para la toma de decisiones sobre el otorgamiento de créditos o el cálculo de pólizas, respectivamente, ya que permiten determinar qué características relativas a los usuarios son las que conllevan mayor o menor riesgo, siendo posible segmentar usuarios en función de estas características.

Ejemplo: Solicitudes de aprobación o no de un crédito hipotecario



Cliente	Ingreso	Historial	Crédito Hipotecario
Si	Bajo	Bueno	Aprobado
No	Alto	Malo	Rechazado
No	Bajo	Bueno	Aprobado
No	Alto	Bueno	Rechazado
Si	Bajo	Malo	Rechazado

Ejemplo: Solicitudes de aprobación o no de un crédito hipotecario



Partes del árbol



Ejemplo: Solicitudes de aprobación o no de un crédito hipotecario



Nodos internos del árbol



Ejemplo: Solicitudes de aprobación o no de un crédito hipotecario



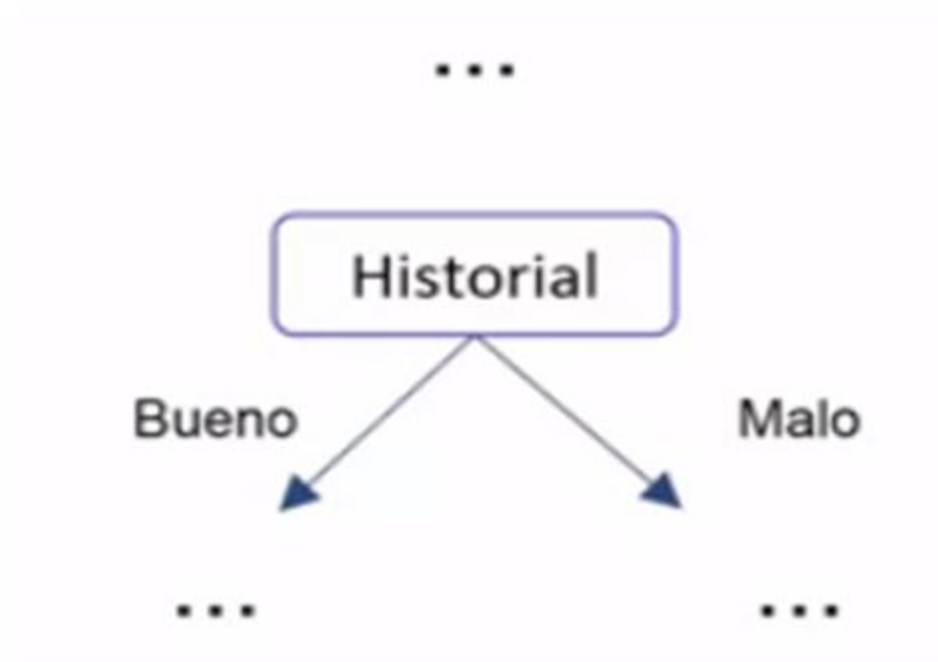
Link de cada nodo



Ejemplo: Solicitudes de aprobación o no de un crédito hipotecario



Atributos



Ejemplo: Solicitudes de aprobación o no de un crédito hipotecario



Nodos hojas



Ejemplo: Solicitudes de aprobación o no de un crédito hipotecario



Clasificación de un registro



Recorriendo el árbol - Ejemplo

¿Cuál es la clasificación para un crédito del cliente con buen historial y alto ingreso?



Recorriendo el árbol - Ejemplo

¿Cuál es la clasificación para un crédito del cliente con buen historial y alto ingreso?



Recorriendo el árbol - Ejemplo

¿Cuál es la clasificación para un crédito del cliente con buen historial y alto ingreso?



¿Cómo hacemos que se construya solo el árbol de decisiones?




Conjunto de datos

Cliente	Ingreso	Historial	Credito Hipotecario
Si	Bajo	Bueno	Aprobado
No	Alto	Malo	Rechazado
No	Bajo	Bueno	Aprobado
Si	Alto	Bueno	Aprobado
No	Alto	Bueno	Rechazado
Si	Bajo	Malo	Rechazado

¿Cómo hacemos que se construya solo el árbol de decisiones?



Selección del nodo raíz

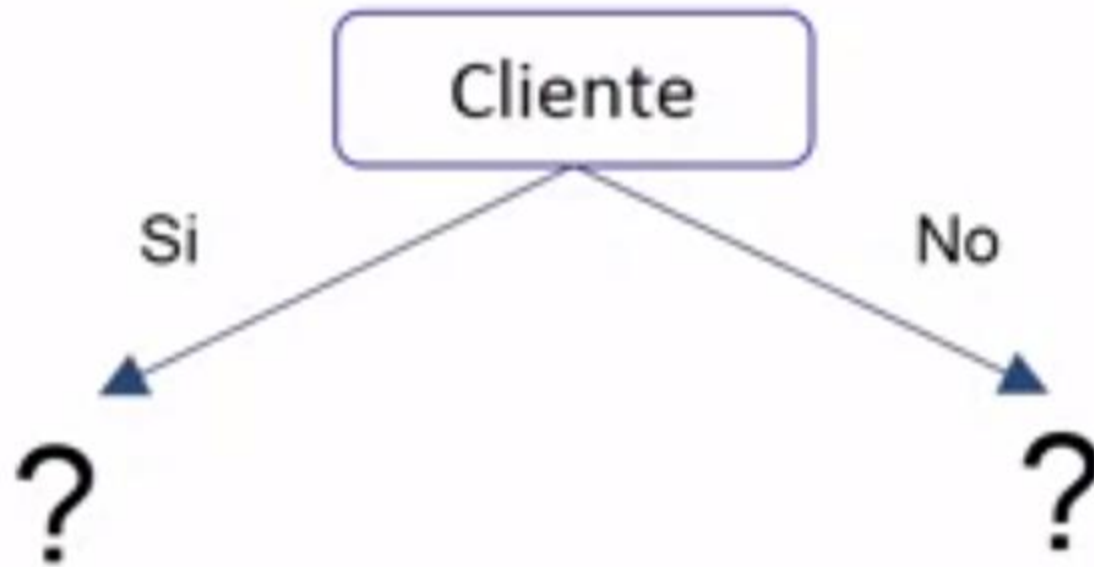


Cliente	Ingreso	Historial	Credito Hipotecario
Si	Bajo	Bueno	Aprobado
No	Alto	Malo	Rechazado
No	Bajo	Bueno	Aprobado
Si	Alto	Bueno	Aprobado
No	Alto	Bueno	Rechazado
Si	Bajo	Malo	Rechazado

¿Cómo hacemos que se construya solo el árbol de decisiones?



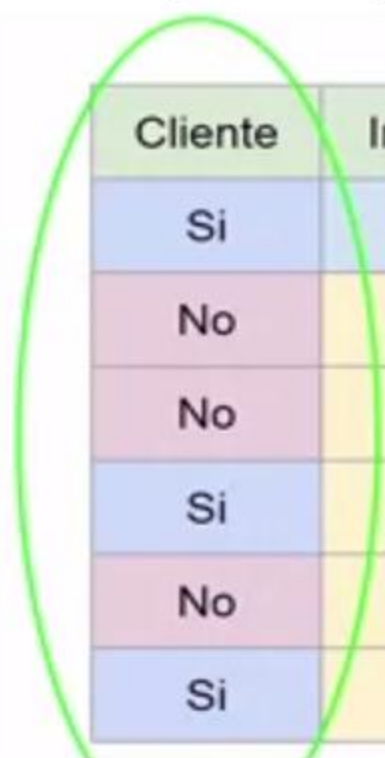
Colocando el nodo raíz



¿Cómo hacemos que se construya solo el árbol de decisiones?



Repitiendo el proceso para la siguiente variable



Cliente	Ingreso	Historial	Credito Hipotecario
Si	Bajo	Bueno	Aprobado
No	Alto	Malo	Rechazado
No	Bajo	Bueno	Aprobado
Si	Alto	Bueno	Aprobado
No	Alto	Bueno	Rechazado
Si	Bajo	Malo	Rechazado

¿Cómo hacemos que se construya solo el árbol de decisiones?

Separar filas de la base de datos

Cliente	Ingreso	Historial	Credito Hipotecario
Si	Bajo	Bueno	Aprobado
No	Alto	Malo	Rechazado
No	Bajo	Bueno	Aprobado
Si	Alto	Bueno	Aprobado
No	Alto	Bueno	Rechazado
Si	Bajo	Malo	Rechazado

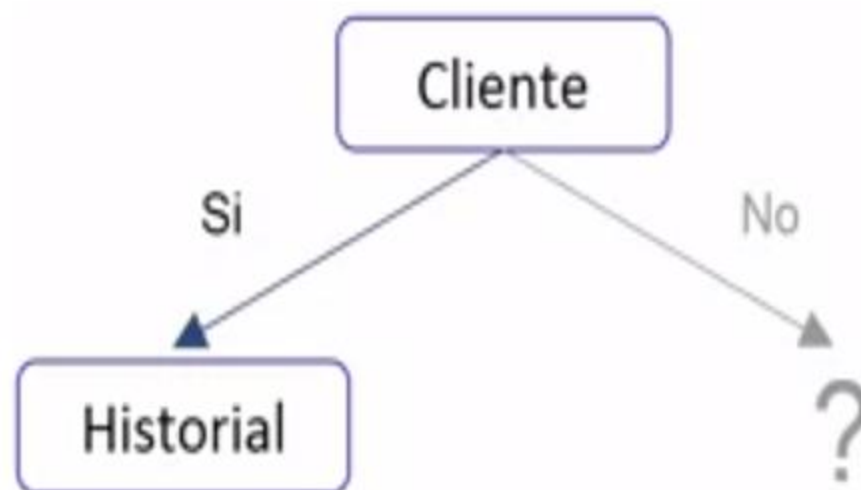


¿Cómo hacemos que se construya solo el árbol de decisiones?



Repitiendo el proceso para la siguiente variable


Ingreso	Historial	Credito Hipotecario
Bajo	Bueno	Aprobado
Alto	Bueno	Aprobado
Bajo	Malo	Rechazado



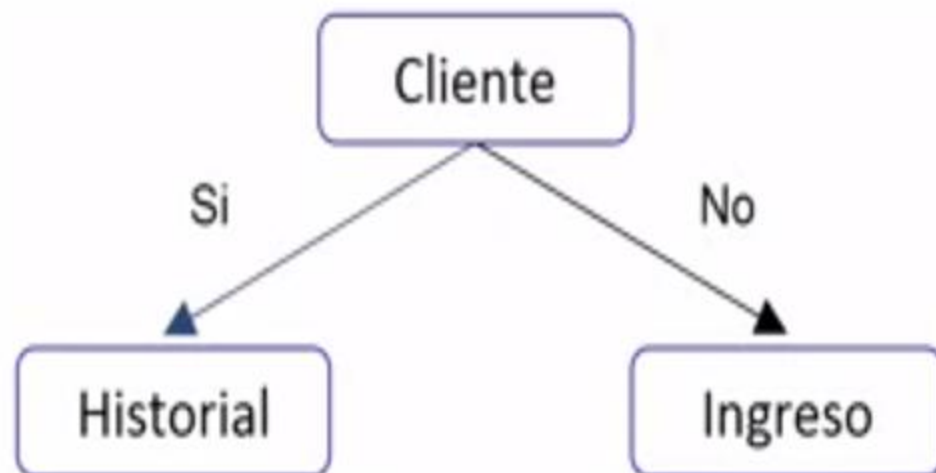
¿Cómo hacemos que se construya solo el árbol de decisiones?



Repitiendo el proceso para la siguiente variable



Ingreso	Historial	Credito Hipotecario
Alto	Malo	Rechazado
Bajo	Bueno	Aprobado
Alto	Bueno	Rechazado



¿Cómo hacemos que se construya solo el árbol de decisiones?



Resultado final



Recursión general del algoritmo contiene los siguientes pasos

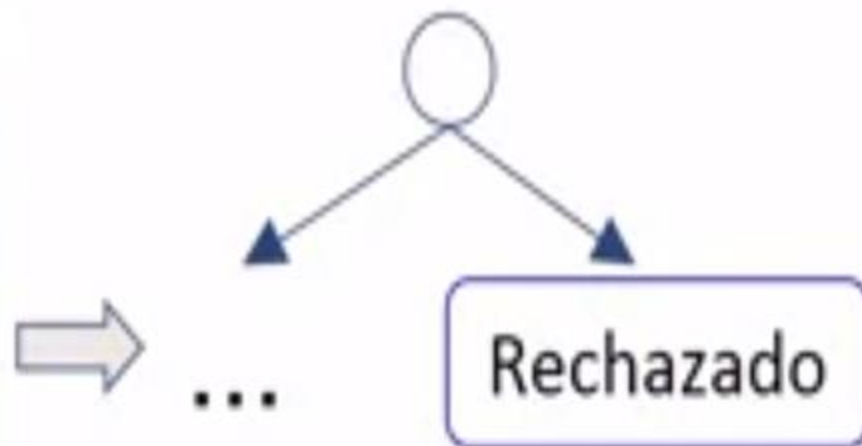


- Tomar el mejor atributo
- Usar este atributo como nodo raíz
- Dividir el set de entrenamiento de acuerdo a este atributo y para cada rama resultante continuar la construcción del árbol recursivamente

Criterios de detención

- ¿Pertenece todos los registros a la misma clase?
- Si → retorna un nodo hoja con la clase respectiva

Ingreso	Historial	Credito Hipotecario
Bajo	Bueno	Rechazado
Bajo	Malo	Rechazado
Alto	Malo	Rechazado



Criterios de detención

- Tienen todos los registros el mismo valor para todos los atributos que determinan su clase
- Si → retorna un nodo hoja con la clase más común



Comencemos con nuestra práctica

