

# **Módulo: ADM1812 - Big Data y Business Analytics - (A50)**

## **Actividad: Actividad de desarrollo 4. Práctica con el software RapidMiner**

**Nombre: Roberto Mora Balderas**

**Asesor: José Carlos Soto Monterrubio**

**Fecha: 16 de junio de 2023**

## Tipo de actividad

Individual

## Objetivo de la actividad

Interpretar los resultados de un árbol de decisión y su respectiva matriz de confusión.

## Instrucciones

Responde las preguntas que se indican con el material proporcionado:

1. *¿Qué significa un análisis de Churn?*

Proceso donde se examina y se comprende la tasa de pérdida de clientes en un negocio o servicio. Se le conoce también como análisis de rotación o de abandono. El termino “churn” hace referencia al abandono de un producto.

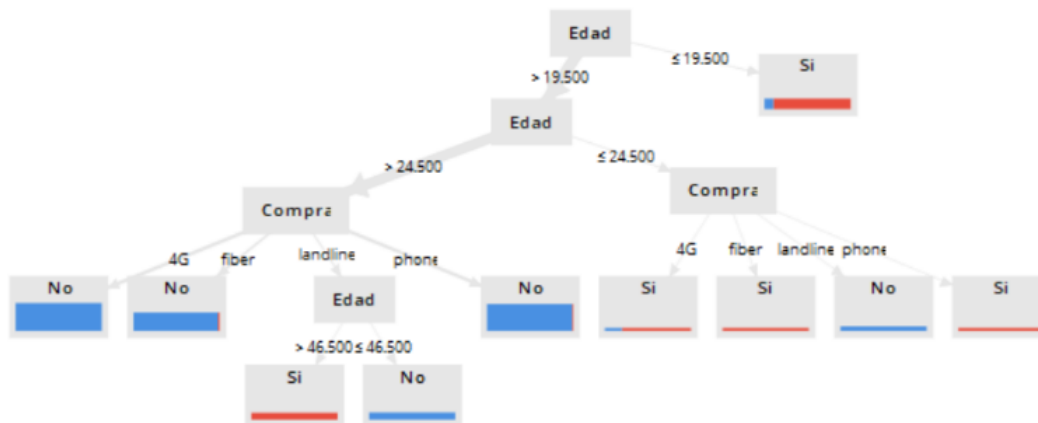
2. *Plantea un escenario y el tipo de data que necesitarías para aplicar un algoritmo de Churn que impacte en los objetivos de tu rol en la organización.*

En el contexto de una empresa petroquímica, donde el objetivo es mejorar la eficiencia y la seguridad operativa de la planta, el uso del algoritmo de Churn puede adaptarse de la siguiente manera:

- i. Definir “Churn” en el contexto de la planta química, ya que no se busca la perdida de clientes, sin embargo, se puede adaptar para identificar eventos de falla o de interrupción inesperada en los equipos y procesos de la planta.
- ii. Recopilación de datos relevantes: Sería necesario recolectar datos relacionados con el funcionamiento de los equipos, como datos de sensores, datos de mantenimiento y cualquier otra información relevante para el monitoreo de la planta.
- iii. El objetivo del modelo sería predecir la probabilidad de que un equipo o proceso específico experimente un evento de Churn (falla o interrupción no planificada).
- iv. Finalmente, con esta data podríamos tener la formula churn de la siguiente manera.

v. 
$$\text{Tasa de Falla} = \frac{\text{Numero de equipos fallando}}{\text{Numero de equipos sin fallas en determinado tiempo}}$$

3. La siguiente imagen es un árbol de decisión elaborada con la data del video demo de Churn durante la unidad; por lo que es un análisis de abandono para clientes de una compañía telefónica, tomando en consideración esto responde:



- a) ¿Cuál es la variable que reduce al mínimo la entropía y nos permite clasificar la mayor parte de los casos de abandono de clientes?

La edad, se observa como es la primera variable que abre al árbol, e incluso sigue siendo la segunda que sigue abriendo el árbol.

- b) ¿Qué sucede con un cliente menor de 19.5 años si sigues el análisis a través del árbol de decisión?

De acuerdo con el árbol de decisiones si el cliente es menor de 19.5 años abandonara el servicio telefónico.

c) ¿Qué predicción tiene un cliente que tiene 30 años y su compra es “phone”?

De acuerdo con el árbol de decisiones si el cliente tiene 30 años con el servicio de “phone” no abandonara el servicio.

4. A partir de la siguiente matriz de confusión responde lo que se indica.

accuracy: 98.88%

	true No	true Si	class precision
pred. No	72	1	98.63%
pred. Si	0	16	100.00%
class recall	100.00%	94.12%	

a) ¿Cuántos falsos positivos surgen en el modelo?

Ninguno.

b) ¿Cuántos falsos negativos surgen en el modelo?

1 falso negativo.

c) ¿Cuál es el total de “verdaderos sí” de la data?

17 verdaderos sí.

d) ¿Cuántos “verdaderos sí” fueron captados correctamente en la predicción?

16 verdaderos sí, fueron captados correctamente.

## Referencias

- Anáhuac Online. (2019). *El árbol de decisión*. [Contenido creado para Anáhuac Online].
- Anáhuac Online. (2019). *Modelos y técnicas para la aplicación del Business Analytics*. [Contenido creado para Anáhuac Online]. Anáhuac Online. (2019). Demo de Business Analytics con el software RapidMiner. [Contenido creado para Anáhuac Online].
- Anáhuac Online. (2019). Matriz de confusión. [Contenido creado para Anáhuac Online].
- Estrada, R. [Anáhuac Online EDX.],. (2020, mayo 16). Matriz de confusión RAFA EDX [video]. [YouTube]. <https://acortar.link/zVkKpy>