

MACHINE LEARNING IMMERSION

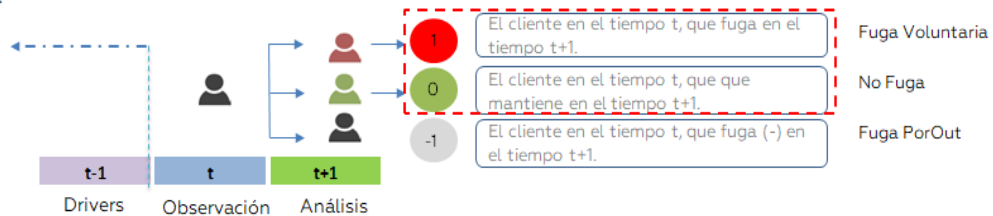
Se entrega el siguiente set de datos de clientes de una entidad financiera, Churn-arboles.csv, el cual es un dataset que cuenta con las principales características del cliente.

Esquema metodológico de la fuga de clientes

Cosechas o periodos de estudio del Churn

Ene-19	Feb-19	Mar-19	Abr-19	May-19	Jun-19	Jul-19
t-3	t-2	t-1	t	t+1	t+2	
	t-3	t-2	t-1	t	t+1	t+2
		t-3	t-2	t-1	t	t+1

Donde:



La empresa en el último mes y debido a la coyuntura del COVID – 19 ha incrementado en + 0.8 pp su ratio de churn de clientes, pasando de 10% en promedio de los últimos 6 meses a 22.5% en los últimos 3 siguientes. Como sabemos el costo de adquisición de un nuevo cliente es aproximadamente 6 a 7 veces mayor que retener uno que ya tenemos. Debido a que usted está llevando un programa de especialización en Data Science requieren que los ayude a **identificar a los clientes con alta probabilidad de fugar de la empresa.**

- Las variables que se disponibilizan son:

Variable	Descripción de la variable
ID	ID o identificación única del cliente.
EDAD	Edad del cliente o persona.
SEXO	Género del cliente o persona.
CIVIL	Estado civil del cliente o persona.
HIJOS	Número de hijos del cliente o persona.
INGRESO	Ingreso del cliente o persona.
AUTO	Tenencia o no de un vehículo por el cliente.
CHURN	Estado de fuga de la empresa. {0: No ; 1:Si}

Entregables del Laboratorio:

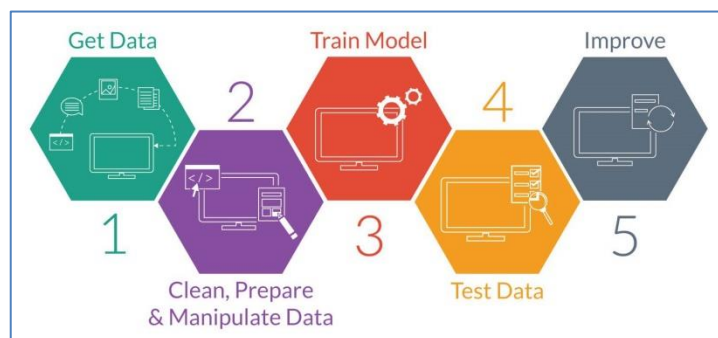
1.- Definir el problema de la naturaleza que se tiene a continuación, además de los objetivos de negocio bien definidos.

2.- ¿Qué tipo de variables se utilizan en el problema de negocio?

3.- Construya mediante Machine Learning una solución analítica para determinar que clientes se irán de la empresa en los próximos meses.

4.- De los drivers o features seleccionados cuáles de éstos son los más importantes para determinar la probabilidad de irse de la empresa.

5.- ¿Cuáles son las principales métricas de negocio o ratios esperados para la gestión con la solución analítica que estamos construyendo?



6.- Con el algoritmo de Machine Learning entrenado, scorear o puntuar la data o información en producción que debería ir a campañas.

