

## Working Analyst

Desarrolle un algoritmo de clasificación para tratar de realizar predicciones dado un DataSet (El algoritmo puede ser cualquiera de los vistos en clase).

Descripción de los datos: El archivo tiene nombre “Datos\_Aseguradora.csv” y contiene las variables

- **Identifiant:** Identificador de los clientes en la base de datos (ID)
- **ft\_2\_categ:** Año en que se observo la póliza de seguro
- **EXPO:** tiempo asegurado en la compañía en el año (Seguro de año completo EXPO=1, 6 meses EXPO=0.5)
- **superficie:** Tamaño, en metros cuadrados, del edificio asegurado
- **Insee:** Código geográfico Frances de donde se ubica el edificio asegurado
- **target:** Es binaria, 0 si no se reclamo el seguro 1 si se reclamo al menos una vez
- **other ft\_i\_categ:** Variables anonimizadas (pro privacidad de clientes) referentes a los edificios asegurados (categóricas)

Realice los siguientes pasos:

- Defina cual será su variable objetivo
- Separe sus datos en 70% para entrenamiento y 30% para validación
- Realice un reescalamiento de los datos (alguno de los usados en clase)
- Convierta las variables categóricas a numéricas con alguno de los metodos vistos en clase (Dummificación, OneHotEncoding, etc)
- Entrene su algoritmo

Compare sus datos obtenidos con los datos reales El entregable deberá ser un archivo .ipynb que contenga:

1. Exploración de los datos
2. Pasos del algoritmo
3. Evaluación e la predicción
4. Gráfica comparando los datos reales con los predichos por el modelo

Conclusiones Extra:

- ¿Por qué utilizar el modelo que desarrolló y no otro?
- ¿En caso de haber tenido que escoger otro, que modelo usaría?