

# Campaña de Márqueting Directo para Banca Minorista.

## Introducción

Una institución de banca minorista desea revisar su estrategia de márqueting directo. Ésta consiste en diversas campañas dirigidas a segmentos específicos de clientes para conseguir un objetivo concreto. Dichas campañas están centralizadas en un "contact center", con el que se contacta al cliente por diferentes canales, muy a menudo el teléfono. Se dispondrá de una serie de datos relativos a las campañas de ventas, que serán analizados para incrementar la eficacia de las mismas.

## Objetivos del Proyecto

### 1. ¿ Cuales son los objetivos del negocio ?

La tecnología de "machine learning" permite mejorar el márqueting utilizando el enfoque de **maximizar el valor del tiempo de permanencia** del cliente mediante la evaluación de los datos disponibles. Se desea, pues, **establecer relaciones más estrechas y duraderas** con el cliente, tomando en consideración la demanda del mercado. Específicamente, el objetivo es **aumentar la contratación de depósitos a plazo** en las campañas de márqueting directo. Por último, el proyecto podría optimizar el número de llamadas, puesto que se obtendría la probabilidad que un cliente determinado acepte un depósito a plazo fijo. Así pues, se podrían **centrar los esfuerzos en los clientes con mayor probabilidad** de éxito.

### 2. ¿ Que decisiones o procesos específicos quereis mejorar o automatizar con ML ?

En particular, en el presente proyecto se desea **predecir el número de éxitos o fracasos de las llamadas** de telemárqueting para la venta de los depósitos a plazo. Adicionalmente se podrá **optimizar el número de llamadas** a realizar para conseguir dicho objetivo.

### 3. ¿ Se podría resolver el problema de manera no automatizada ?

Si bien el objetivo puede plantearse mediante herramientas estadísticas, la gran cantidad de datos disponibles hace aconsejable el uso de "machine learning". Por otro lado, los algoritmos "ML" permitirían agilizar la toma de decisiones relativas a las campañas de márqueting.

## Metodología Propuesta

Dado que se desea predecir el "éxito" o "fracaso" de una venta, se escogerá un modelo de clasificación. La "**regresión logística**" no solo predice (clasifica) el resultado de la venta, sino que además nos facilita la **probabilidad de dicho éxito**. Se escoge pues dicho modelo.

## Datos Disponibles

Se dispone de una serie estadística de datos de una longitud de 11.162 registros de ventas anteriores de depósitos a plazo.

Las primeras 4 columnas contienen datos demográficos, la edad, el empleo, el estado civil y el nivel educativo del cliente.

Las siguientes 4 columnas muestran datos bancarios del cliente. A saber, si está en números rojos, el saldo medio en su cuenta, si posee una hipoteca y si posee un crédito personal.

Los siguientes 7 campos se refieren a las propias campañas de telemárqueting, el tipo de contacto, día de la semana y mes del contacto, la duración de la llamada en segundos y otras relacionadas.

Las dos últimas se refieren al éxito o fracaso de la llamada. La primera, "outcome", se refiere a si el cliente contrató un depósito a plazo en anteriores campañas. La segunda es nuestro objetivo, nuestra predicción, esto es, si el cliente ha contratado un depósito a plazo fijo en la presente campaña.

## Métrica de éxito del proyecto

Para los modelos de clasificación pueden utilizarse dos métodos el análisis LIFT y la gráfica ROC.

El análisis LIFT consiste en dividir la probabilidad de éxito predicha (la probabilidad de contratar el depósito) en deciles, esto es en 10 grupos: 0.0-0.1, 0.1-0.2, ... , 0.9-1.0.

Dichos valores formarán el eje horizontal de un gráfico LIFT. Para el eje vertical se dividirán el número de éxitos en cada grupo, por el número de clientes de dicho grupo. En los grupos con poca probabilidad deberían encontrarse pocos éxitos (pocos contratos), en los grupos con mucha probabilidad deberían contarse muchos éxitos. Lo contrario indicaría que la capacidad de predicción del método es baja. La puntuación LIFT se define como:

$$\text{lift} = (\text{probabilidad predicha}) / (\text{probabilidad media})$$

El parámetro permite centrar los esfuerzos en los grupos con un LIFT mayor, puesto que identifica los grupos con el mayor número de éxitos predichos.

La gráfica ROC se basa en la matriz de confusión, de la que se obtienen los parámetros:

$\text{TPR} = \text{TP} / \text{P}$ , positivos auténticos entre total positivos.

$\text{FPR} = \text{FP} / \text{N}$ , falsos positivos entre total negativos.

Si TPR es igual a FPR estaríamos en el así llamado "clasificador aleatorio" y nuestro modelo tendría poca capacidad de predicción. Cuanto mayor sea TPR respecto a FPR tanto mejor será la predicción del modelo y viceversa.

## Responsabilidades Éticas y Sociales

De cara a los principios éticos aplicados a la IA, se definen cuatro roles: Gobiernos, empresas, desarrolladores y usuarios.

En calidad de **desarrolladores** tendremos en cuenta la privacidad, seguridad, equidad y transparencia durante el diseño e implementación del proyecto.

En cuanto a la **privacidad**, en los datos a considerar aparecen los nombres, emails y direcciones de los usuarios que serán tratados como privados.

En cuanto a la **seguridad**, el proyecto no trata con materias relacionadas con la seguridad de personas o equipos, a diferencia de aplicaciones como la conducción autónoma o aplicaciones industriales.

En cuanto a la **equidad**, aparecen en los datos la edad, el empleo, el estado civil y el nivel educativo. Deberán por tanto evitarse los sesgos relacionados con tales datos.

En cuanto a la **transparencia**, se utiliza como modelo la regresión lógica, siendo éste un modelo ampliamente utilizado. Se comprobará su capacidad de predicción mediante las métricas ROC y LIFT.