

Módulo de Machine Learning: Caso Final

Una empresa internacional relacionada con el sector de las telecomunicaciones quiere afianzarse nacionalmente. Para ello, quiere fidelizar a sus clientes aumentando el número de compras de estos llevando a cabo campañas de crossselling.

El cross-selling o venta cruzada consiste en sugerir al cliente productos complementarios al que acaba de comprar con el objetivo de brindar una experiencia más completa y aumentar el número de ventas. En este caso, los clientes activos pagan mensualmente una cuota por servicios de telecomunicaciones, por lo que reciben anuncios/sugerencias después de cada pago.



El objetivo del proyecto es realizar un modelo que prediga qué clientes son más propensos a realizar otra compra además del pago por los servicios habituales. El equipo de la empresa internacional no sabe muy bien cómo abordar el problema, por lo que te contratan para ayudarles a desarrollar el proyecto. Te pueden proporcionar la información de su base de datos, en el que reúnen la información de los clientes activos (todos los meses pagan por un servicio). Cuentan con un total de 20.000 muestras con 34 variables relativas a clientes tomadas mensualmente. A cada muestra, se le ha asignado una etiqueta que confirma si el cliente ha comprado o no un producto ese mes.

1. Información de los datos:

El dataset cuenta con **34 variables descriptivas X**. El número total de muestras es de 20 mil clientes.

Variables independientes X:

1. MB_TOTALES: mb totales consumidos por el cliente
2. FACTURACION_TOTAL_IMPUESTOS: facturación del cliente incluyendo impuestos
3. FACTURACION_CUOTA: facturación sólo de la cuota de los servicios
4. SEGUNDOS_LLAMADA_RED_PROPIA
5. SEGUNDOS_LLAMADA_SALIENTE_RED_PROPIA
6. SEGUNDOS_LLAMADA_ENTRANTE_RED_PROPIA

7. MB_MENSUALES
8. EDAD: edad aproximada del cliente con un rango de error +/-5
9. NUM_LLAMADAS_RED_PROPIA
10. FACTURACION_TOTAL_SIN_IMPUESTOS
11. NUM_LINEAS_PRIN_POSPAGO: líneas principales de contrato
12. NUM_LINEAS_POSPAGO
13. NUM_LINEAS_FIJAS
14. NUM_DIAS_CONVERGENTE: días que el cliente lleva con un fijo y un móvil
15. NUM_LINEAS_FIJAS_POSPAGO
16. NUM_SERVICIOS_POSPAGO
17. DIAS_PRIMERA_CUENTA: días desde primera cuenta en la empresa
18. NUM_CANCELACIONES
19. LLAMADAS_MOVILES
20. NUM_LINEAS_POSPAGO_TOTAL
21. TERMINO_FACTURACION
22. NUM_DESACTIVACIONES_FIJAS_POSPAGO
23. NUM_DESACTIVACIONES_FIJAS
24. NUM_LINEAS_TECNOLOGIA_DESCONOCIDA
25. NUM_DIAS_ACTIVO
26. MINUTOS_LLAMADAS_MENSUALES
27. NUM_DIAS_BUNDLE: número de días totales en el que el cliente ha tenido error en algún servicio
28. KPI_IMPUESTOS: impuestos de la factura principal del cliente
29. NUM_DESACTIVACIONES
30. PORCENTAJE_SEGUNDOS_RED_PROPIA
31. SERVICIOS_TOTALES_MARCA
32. SERVICIOS_TOTALES_POSPAGO
33. MINUTOS_LLAMADAS_MOVIL
34. PENETRACION_FIJO_ZONA: porcentaje de personas en la ciudad del cliente que tienen un fijo

Variable independiente Y:

1. TARGET: variable que mide si el cliente adquiere un producto adicional (1) o no (0) .

Variable temporal MES:

Esta variable recoge el mes en que se ha tomado la muestra. Ej: si el mes de datos es 201906 y TARGET es 1, significa que ese cliente el mes de junio adquirió otro producto. Si la target es 0, significa que ese cliente no adquirió otro producto el mes de Junio.

2. Tareas para completar el proyecto:

- a. Análisis exploratorio del dataset proporcionado (relación entre variables, tratamiento de nulos, outliers, análisis de correlación...). (1.5 puntos)
- b. Preparación de los datos para el modelado (selección de variables de entrenamiento, estandarización, división del dataset...). (1.5 puntos)
- c. Comparación del rendimiento de varios modelos:
 - Regresión Logística
 - Modelo ensamblado
 - Red Neuronal (MLP)

Se valorarán las técnicas aplicadas para la selección de hiperparámetros, así como las métricas elegidas para la evaluación del modelo. (5 puntos)

- d. Se sospecha que la segmentación de clientes mejoraría el rendimiento del modelo predictivo, por lo que se pide dividir el dataset en dos clústeres y entrenar y ajustar el mejor modelo del apartado anterior para cada uno de los dos grupos de clientes. (2 puntos)
- e. (PUNTO EXTRA) Tratamiento y análisis de la columna Mes. (1 punto)

3. Consideraciones de la práctica:

- La práctica se resolverá en un Notebook de Python que se entregará del siguiente modo:
 - Debe llamarse ML2024_Nombre_Apellidos.
 - Se entregará en formato .ipynb.
 - Todos los kernels del Notebook deben estar ejecutados.
 - La entrega se realizará en el apartado habilitado de BlackBoard (Caso Final Grupo 1, dentro de la carpeta de recursos del grupo 1).
- Todas las decisiones deben aparecer justificadas y los resultados comentados.
- El Notebook debe estar estructurado y ordenado.
- Al final, se deben comentar las conclusiones finales del ejercicio.