FINAL PRACTICE

Machine learning

Brief description

A telecommunications company aims to retain customers through **cross-selling** campaigns, suggesting products after their monthly service payment. They hire you to develop a predictive model to identify customers likely to make additional purchases. They have 20,000 samples and 34 monthly variables.

Sergio Hervás Aragón
Sergiohervas9@gmail.com

An international company related to the telecommunications sector wants to establish itself nationally. To do this, it wants to build customer loyalty by increasing the number of purchases by these customers by carrying out cross-selling campaigns. Cross-selling consists of suggesting complementary products to the customer to the one they have just purchased with the aim of providing a more complete experience and increasing the number of sales. In this case, active customers pay a monthly fee for telecommunications services, so they receive advertisements/suggestions after each payment.

The goal of the project is to develop a model that predicts which customers are more likely to make an additional purchase besides their regular service payment. The international company's team is unsure how to approach the problem, so they hire you to help develop the project. They can provide you with their database, which contains information on active customers (who pay for a service every month). They have a total of 20,000 samples with 34 customer-related variables collected monthly. Each sample is labeled to indicate whether the customer made a purchase that month.

Variable Information

Each dataset must include the following columns:

Independent variables X:

- MB_TOTALES: Total MB consumed by the client
- FACTURACION_TOTAL_IMPUESTOS: Customer billing including taxes
- FACTURACION_CUOTA: billing only for the service fee
- SEGUNDOS_LLAMADA_RED_PROPIA
- SEGUNDOS_LLAMADA_SALIENTE_RED_PROPIA
- SEGUNDOS_LLAMADA_ENTRANTE_RED_PROPIA
- MB_MENSUALES
- EDAD: Approximate age of the client with an error range of +-5
- NUM_LLAMADAS_RED_PROPIA 10. FACTURACION_TOTAL_SIN_IMPUESTOS
- NUM_LINEAS_PRIN_POSPAGO: main lines of contract
- NUM_LINEAS_POSPAGO 1
- NUM_LINEAS_FIJAS
- NUM_DIAS_CONVERGENTE: Days that the customer has had a landline and a mobile pone.
- NUM_LINEAS_FIJAS_POSPAGO
- NUM_SERVICIOS_POSPAGO
- DIAS_PRIMERA_CUENTA: days since first account in the company
- NUM_CANCELACIONES
- LLAMADAS_MOVILES
- NUM_LINEAS_POSPAGO_TOTAL
- TERMINO_FACTURACION
- NUM_DESACTIVACIONES_FIJAS_POSPAGO
- NUM_DESACTIVACIONES_FIJAS
- NUM_LINEAS_TECNOLOGIA_DESCONOCIDA
- NUM_DIAS_ACTIVO
- MINUTOS_LLAMADAS_MENSUALES
- NUM_DIAS_BUNDLE: Total number of days in which the client has had an error in some service.
- KPI_IMPUESTOS: customer main invoice taxes
- NUM_DESACTIVACIONES
- PORCENTAJE_SEGUNDOS_RED_PROPIA
- SERVICIOS_TOTALES_MARCA
- SERVICIOS_TOTALES_POSPAGO
- MINUTOS_LLAMADAS_MOVIL

- PENETRACION_FIJO_ZONA: percentage of people in the client's city who have a landline.

Independent variable Y:

- TARGET: Variable that measures whether the customer purchases an additional product (1) or not (0).

OBJECTIVES

With the given datasets and variables given, you are required to:

1. Exploratory analysis of the dataset provided (relationship between variables, treatment of nulls, outliers, correlation analysis...).
2. Data preparation for modeling (selection of training variables, standardization, dataset division, etc.).
3. Comparison of the performance of various models:
    - Logistic regression
    - Assembled model
    - Neural Network (MLP)
4. It is suspected that customer segmentation would improve the performance of the predictive model, so it is requested to divide the dataset into two clusters and train and adjust the best model from the previous section for each of the two groups of customers.
5. Treatment and analysis of the Month column.