

MBD – Estadística – Práctica II (RLOG)

Descripción

Uno de los elementos de la estrategia de Márketing de cualquier entidad financiera es cómo informar a los usuarios de sus nuevos productos y servicios en unas condiciones que maximicen las opciones de adquisición. El desarrollo y aplicación de la segmentación de clientes es una condición necesaria para competir adecuadamente en los diferentes ámbitos que la entidad financiera haya definido como mercado objetivo.

El telemarketing es una de las múltiples vías empleadas por las entidades financieras para ofrecer sus productos. Estas llamadas tienen unos costes de personal y de consumo que hacen primordial una buena selección de los individuos a los cuáles dirigir la venta. La entidad bancaria Just-Your-Bank (JYK) tiene el propósito de reducir dichos costes reduciendo el número de llamadas realizadas a aquellos clientes con pocas opciones de contratar el producto ofrecido.

El objetivo de esta práctica es predecir la probabilidad que un cliente al cual se realiza la llamada acepte el producto que se le ofrece.

Datos

Se entregarán dos conjuntos de datos con registros de clientes de JYK a los que se les realizó una llamada para vender un producto concreto:

1. Datos de entrenamiento. Contendrán la variable respuesta (contrató o no contrató el producto)
2. Datos test. No contendrán la variable respuesta y ésta deberá predecirse basándose en los datos históricos del conjunto de entrenamiento.

Las variables comunes que contienen ambos juegos son:

Características del cliente

id: identificador del cliente

age: edad

job: tipo de trabajo (admin., blue-collar, entrepreneur, housemaid, management, retired, self-employed, services, student, technician, unemployed, unknown)

marital: estado civil (divorced, married, single, unknown)

education: nivel de estudios (basic.4y, basic.6y, basic.9y, high.school, illiterate, professional.course, university.degree, unknown)

default: ¿es moroso? (no,yes,unknown)

housing: ¿tiene hipoteca? (no,yes,unknown)

loan: ¿tiene un préstamo personal? (no,yes,unknown)

Características de la llamada

contact: tipo de teléfono (cellular, telephone)

month: mes

day_of_week: día de la semana (mon, tue, wed, thu, fri)

Otros atributos

campaign: número de contactos realizados esta campaña para este cliente (incluyendo el actual)

pdays: número de días que han pasado desde que el cliente fue contactado por última vez para una campaña previa (999 significa que no fue contactado previamente)

previous: número de llamadas realizadas a este cliente antes de esta campaña

poutcome: resultado de la anterior campaña (failure, nonexistent, success)

Indicadores del contexto social y económico

emp.var.rate: indicador de la tasa de empleo (cuatrimestral)

cons.price.idx: IPC (mensual)

cons.conf.idx: Índice de confianza del consumidor (mensual)

euribor3m: euribor a 3 meses (diario)

nr.employed: número de empleados (cuadrimestral)

Variable respuesta (sólo en el juego de entrenamiento):

Y: ¿Se suscribió el cliente al depósito? (yes,no)

Evaluación

Se basará en el AUC (área bajo la curva ROC) del modelo aplicado sobre la prueba test.

$$\int_{-\infty}^{\infty} TPR(T) \cdot FPR'(T) \cdot dT$$

Donde:

TPR es la proporción de verdaderos positivos

FPR es la proporción de falsos positivos

Se obtendrá una puntuación más alta cuanto mayor sea este estadístico. Adicionalmente se valorará:

1. La parsimonia del modelo
2. Los criterios de selección del modelo
3. La presentación de los resultados

Entrega

La fecha límite para realizar la entrega será el día 11/12/2022 a través del campus.

La documentación a entregar será:

1. Una descripción del trabajo realizado. El documento debe incluir la expresión del modelo final y el proceso de selección del mismo. Además se debe especificar si se creó alguna variable derivada o si se realizó alguna transformación sobre las ya existentes Extensión: 3 o 4 páginas. Formato: *.docx/pdf*
2. Fichero de texto con la probabilidad predicha para el juego test en dos columnas separadas por una coma: la primera debe contener el identificador del cliente y la segunda la probabilidad de contratación del depósito. Formato: *.txt*. Es muy importante que este fichero se llame exactamente "p2.txt"
3. El código utilizado para realizar esta parte con comentarios explicativos. Formato: *.R*