

cognodata,

EXAMEN DE MODELACIÓN

.2.

Tenemos la siguiente BBDD en una compañía:

- Tabla "CLIENTES_STATUS" con las siguientes columnas:

- * Id Cliente
- * MES
- * Variables de status de cliente
 - * Nº impagos
 - * % descuento a aplicar en ofertas
 - * ...

- Tabla "CLIENTES" con las siguientes columnas:

- * Id Cliente
- * Fecha Alta
- * Sexo
- * CP
- * Dirección
- * Edad
- *

- Tabla "FAMILIAS" con las siguientes columnas:

- * Id Familia
- * Id Producto

- Tabla "VENTAS" con las siguientes columnas:

- * Id Producto
- * Mes
- * Ventas Totales
- * Ventas al estado
- * ...

...

* Required

1. Escriba su nombre completo (nombre y apellidos) *

El texto "Su respuesta es anónima" indica que no se requiere autenticación

Jose Angel Velasco

2. Proponga información relevante y/o tablas adicionales que pudieran ser de ayuda a la hora de predecir la fuga de clientes de la compañía. Cada propuesta debe justificarse apropiadamente.

Datos personales del cliente (tablas FAMILIAS, CLIENTES, y alguna adicional del ERP como estudios, profesión, renta del municipio donde viven, etc.), Datos de contratación (Tabla CLIENTES, ,CLIENTES STATUS), Datos de uso (tabla VENTAS, y alguna adicional referido al uso diario del servicio), Datos de facturación (tabla ventas, y alguna adicional referida a los pagos del cliente), Datos de otras fugas (duración del contrato del cliente fugado, localidad, nivel de estudios, profesión, tipo de producto contratado, número de

3. Defina variable sintética (o variable transformada) y proponga ejemplos válidos que ayuden a la predicción de la fuga de clientes

Variable sintetica: variable creada en base a otras variables existentes mediante la combinación lineal o no lineal. Por ejemplo: una variable sintética en este problema podrir ser un modelo agregado de el tiempo medio diario del uso del servicio mas el importe medio del coste mensual del servicio sobre el importe medio del mercado para un producto similar mas el tiempo acumulado del contrato del servicio. Clientes con un alto consumo que pagan una cantidad elevada de dinero pueden ser mas propensos a abandonar la compañía por otra que les ofrezca el mismo suministro a un precio más

4. Defina el target (o variable objetivo) "fuga de clientes" en este ejemplo con el mayor detalle posible. ¿Qué algoritmos emplearía para predecir la fuga de clientes? ¿Por qué? Describa el proceso de modelación con una de las técnicas, indicando todos los pasos que considere necesarios para obtener un resultado óptimo.

La variable objetivo en este problema sería una "etiqueta" con "Si" o "No" y sería el resultado de la predicción de fuga del cliente. Emplearía un modelo no-paramétrico que no hiciera grandes hipótesis sobre la estructura subyacente del problema como un ensemble de árboles de decisión (Random Forest) para garantizar un compromiso adecuado entre bias y varianza (flexibilidad para adecuarse al problema y varianza del resultado predicho cuando nuevos clientes sean testados). Primero realizaría un proceso de limpieza de datos (supresión de duplicados y valores nulos, detección de valores fuera de rango y outliers e inconsistencias, filtrado, normalización y codificación de variables categóricas y agregación por cliente. Después determinaría las features a usar para resolver el problema de clasificación mediante análisis de correlación y varianza, así como análisis por PCA (principal component analysis). Una vez identificadas las features a usar como inputs del Random Forest, ejecutaría un entrenamiento por K-fold cross-validation ($k=4$, a priori) combinado con una optimización de los hyper parámetros del modelo mediante un grid search definiendo un conjunto de valores para cada hyperparámetro (número estimadores, profundidad del árbol, mínimo número de

5. Se dispone del modelo ya entrenado, y la compañía lo quiere utilizar para lanzar una campaña de fidelización al 10% de los clientes que presenten mayor propensión a la fuga. Estamos a día 3 de Mayo, y la compañía nos comunica que hasta el día 10 de Mayo no tendremos disponibles los datos correspondientes al mes de Abril.

- Proponga un calendario de explotación del modelo, indicando los meses utilizados para entrenamiento, para testeo y para explotación.
- Defina un sistema de medición de los resultados del modelo en la fase de explotación.
- Defina "grupo de control". ¿Sería necesario crear uno en este caso? En caso afirmativo, describa brevemente cómo lo haría.

- Utilizaría un año de datos para el entrenamiento K-fold ($k=4$) de forma que el algoritmo "ve" todo el año de datos, para la explotación utilizaría los datos del último mes.
- Utilizaría la métrica recall: número casos de fuga detectados/número de casos de fuga posibles
- El grupo de control se refiere a la prueba de otra versión del algoritmo (ya sea por que se ha entrenado con datos diferentes, tiene otros hyperparámetros o es otro

6. La compañía desea, adicionalmente, realizar una segmentación corporativa de sus clientes

¿Qué técnicas emplearía para dicha segmentación?

- d. ¿Qué técnicas emplearía para dicha segmentación?
- b. ¿Qué ejes considera fundamentales para segmentar a los clientes?
- c. Imagínese una segmentación óptima ya ejecutada. Cite la utilidad que podría sacarle la compañía a dicha segmentación.
- d. La compañía nos especifica que quiere un segmento específico para todas las mujeres clientes de la compañía. Describa en 4-5 líneas cómo adaptaría el algoritmo de segmentación para tener dicho segmento.

a) realizará un clustering por K-means++ ya que es un método de clustering no supervisados muy versátil y adecuado dado un conjunto de features.

b) La respuesta nos la daría el resultado del feature engineering o el análisis por PCA pero a priori diría que podría ser el importe medio mensual del servicio y el consumo del servicio medio.

c) El disponer de clusters de clientes agilizaría todas las campañas que la compañía quisiera realizar ayudando también a enfocar el objetivo de dicha campaña en función de

7. ¿Sería posible aprovechar los dos modelos planteados (de predicción de fugas y segmentación de clientes) de forma que uno sirva para mejorar al otro? Indique cómo lo haría.

Claro, el proceso de clustering, sería un paso previo al modelo de predicción de fugas. Se entrenaría un modelo de predicción de fugas por cada cluster de clientes

Submit

Never give out your password. [Report abuse](#)

This content is created by the owner of the form. The data you submit will be sent to the form owner. Microsoft is not responsible for the privacy or security practices of its customers, including those of this form owner. Never give out your password.

Powered by Microsoft Forms | [Privacy and cookies](#) | [Terms of use](#)