

Tarea

Diplomado Inteligencia de Negocios 2021

Considere el problema que enfrenta una entidad financiera que posee altas tasas de fuga voluntaria. Esta institución no tiene claro cuál es el perfil característico que tienen los clientes fugitivos ni cuáles son las razones por las cuales estos se fugan. Dentro de este contexto, le han encomendado a usted dos tareas fundamentales. La primera tarea consiste en desarrollar un modelo predictivo que permita identificar tempranamente cuáles clientes son más propensos a fugarse. La segunda tarea consiste en definir una serie de políticas comerciales que permitan retener a estos potenciales clientes fugitivos.

Adicionalmente, en la Tabla 1 se muestran los beneficios/costos económicos de los aciertos y errores en las predicciones. La decisión comercial luego de haber aplicado el modelo de clasificación es definir a qué clientes cobrarles una tarifa normal y a qué clientes ofrecerles una campaña. En este caso, la realización de una campaña corresponde a la acción comercial de retención, la que de ser aplicada a un cliente que no se hubiese fugado tras el periodo de evaluación, significa un beneficio de -100. Por otra parte, si la acción de retención se aplica sobre un cliente que se hubiese fugado, se produce un beneficio de 1000.

		Resultado	
		Fuga	No Fuga
Predicción	Fuga	1000	-100
	No Fuga	-5000	0

Tabla 1. Matriz de beneficios/costos de las acciones comerciales.

De acuerdo con lo anterior, es necesario responder a las siguientes interrogantes expuestas por el gerente general:

1. ¿Qué hacer para que estos clientes no se fuguen?
2. ¿Cuáles son las acciones comerciales que deberíamos emprender con estos clientes?
3. ¿Se deben aplicar estas medidas a todos los clientes?

Una completa descripción de las variables recolectadas por cada cliente se presenta a continuación:

1. ID: identificador del cliente.
2. Genero: Genero del cliente.
3. Renta: Renta en pesos.
4. Edad: Edad en años.
5. NIV_Educ: Nivel educacional.
6. E_Civil: Estado civil.
7. COD_Ofi: Código de la oficina
8. COD_Com: Código de la comuna.
9. Ciudad: Ciudad de la oficina.
10. D_Marzo: Deuda de Marzo.
11. D_Abril: Deuda de Abril.
12. D_Mayo: Deuda de Mayo.
13. D_Junio: Deuda de Junio.

14. D_Julio: Deuda de Julio.
15. D_Agosto: Deuda de Agosto.
16. D_Septiembre: Deuda de Septiembre.
17. M_Moroso: Meses en Mora.
18. Monto: Monto crédito preaprobado.
19. Seguro: Seguro de gravamen.
20. Fuga: Variable objetivo.

El primer archivo llamado BASEFUGA_GENERAL.csv contiene los datos con la variable dependiente indicando si se fugó o no el cliente. El segundo archivo llamado BASEFUGA_VALIDACION.csv, provee de otro conjunto con 1.200 clientes descritos por las mismas variables, salvo que no se conoce si se fugaran o no.

Resultados Esperados

1. Desarrollo de un modelo predictivo de fugas que maximice el beneficio esperado.
2. Definir el patrón característico de los clientes fugitivos y de los clientes no fugitivos.
3. Sugerencia de al menos cinco acciones comerciales a emprender con los clientes que el modelo indique (se valorará la sensatez comercial y lo efectivas que estas puedan ser para evitar la fuga).
4. Predicción del conjunto BASEFUGA_VALIDACION.csv (la no entrega de estos resultados significa una penalización de 0.5 pts en la nota final).

Para la tarea se les pide:

1. Indicar qué problemas detecta en la base de datos y medidas para corregirlos. Debe indicar el procedimiento que siguió para resolverlos de forma clara y precisa.
2. Realizar un análisis descriptivo de los datos. En este punto deben calcular las principales métricas estadísticas para una mejor descripción de los datos, por ejemplo, estadísticas descriptivas, tablas de frecuencia, etc. Apoye su análisis utilizando gráficos.
3. Transformar datos y generar indicadores: transforme las variables disponibles de forma de encontrar nuevas variables, generadas a partir de las originales, que puedan tener un buen desempeño predictivo a partir de sus conocimientos del negocio.
4. Seleccionar atributos: analice sus variables de la base. ¿Discriminan entre clientes fugados y no fugados?.
5. Modelación de Minería de Datos: **entrene al menos 4 modelos predictivos**. Algunos candidatos son: regresión logística, k-vecinos más cercanos, árboles de decisión y SVM, no obstante, también puede utilizar modelos cuya implementación no hayamos visto en el programa. Describa el proceso de validación y selección del modelo para cada caso. Describa las soluciones encontradas en términos de las variables presentes y coeficientes. Aplique este modelo al conjunto de datos de predicción.

Escriba un reporte en el mismo notebook donde desarrolle los análisis para su jefatura, documentando cada etapa del procedimiento desarrollado para generar esta base de datos y sus recomendaciones.

Pauta reporte

El informe (en notebook) desarrollado debe contener al menos los siguientes puntos:

1. Nombre de los integrantes e índice.
2. Introducción: se debe dar una descripción general del problema a resolver, una breve descripción del contenido del reporte y una clara visualización de objetivos, tanto generales como específicos (todo lo referido a la documentación debe ser desarrollado en markdown).
3. Análisis estadístico de los datos
4. Modelos de Minería de Datos: no olvidar la descripción de parámetros de ajuste respecto a la medida de performance (beneficios esperados).
5. Determinar el mejor modelo de clasificación utilizando como criterio el máximo beneficio y definición del patrón característico de los clientes fugas y no fugas.
6. Políticas comerciales, conclusiones y discusiones.

Condiciones de la tarea:

- Debe ser desarrollada en grupos de 2 a 3 personas.
- El software a utilizar será Python en Jupyter Notebook.
- Se pide un reporte totalmente autocontenido en Jupyter y un archivo formato .csv con la predicción para todas las observaciones del conjunto de BASEFUGA_VALIDACIÓN. El mejor grupo tendrá una bonificación de 0,5 puntos en la nota de esta tarea.
- La fecha de entrega **será la estipulada en u-cursos.**