

Actividad unidad 3

Análisis del perfil crediticio de clientes bancarios

Se te va a proporcionar un fichero CSV con 45000 registros. Estos registros representan clientes de una entidad bancaria, para cada uno de los cuales hay 14 atributos descriptivos. Los atributos están relacionados con aspectos de gestión de préstamos e hipotecas otorgados, o no, a dichos clientes.

Modelo 1

La entidad bancaria quiere poder predecir, de forma automatizada, si se concede el préstamo o no para posibles nuevos solicitantes, y también quiere poder obtener la puntuación del riesgo. Para ello, quieren que implementemos un modelo de aprendizaje, el cual, a partir de los datos proporcionados, pueda ofrecer esa información. Como la entidad bancaria está en el extranjero, la información que nos proporciona está en inglés. A continuación, se describen las variables del dataset y una breve descripción:

Modelo 2

Column	Description	Type
age	Age of the person	Float
gender	Gender of the person	Categorical
education	Highest education level	Categorical
income	Annual income	Float
yrs_exp	Years of employment experience	Integer
ownership	Home ownership status (rent, own, mortgage...)	Categorical
amount	Loan amount requested	Float
purpose	Purpose of the loan	Categorical
int_rate	Loan interest rate	Float
percent_income	Loan amount as a percentage of annual income	Float
cred_hist_length	Length of credit history in years	Float
risk_score	Credit risk score of the person	Integer
previous_loans	Indicator of previous loan defaults	Categorical
loan_status	Loan approval status: 1 = approved; 0 = rejected	Integer

Target del modelo 1

Pasar todas las int a float (provisional por mayoría)

La variable "loan_status" es la que determina si se ha aprobado o rechazado la solicitud del préstamo, y la variable "risk_score" es la que determina el riesgo crediticio del cliente.

Target del modelo 2

En esta actividad tendrás que desarrollar, en un notebook de Jupyter, todos los pasos para llegar a formular los modelos mencionados. Utiliza las librerías vistas hasta ahora, ninguna más; y emplea los pasos y técnicas trabajados, especialmente en la unidad 3, sin perder de vista los conocimientos útiles de las unidades anteriores y que puedan estar relacionados.

Es importante que cuides el aspecto del notebook, y que emplees el formato "markdown" para introducir cada apartado, así como para describir y detallar todas las aclaraciones, explicaciones y conclusiones oportunas.

Tendrás que realizar la carga, limpieza, ingeniería de características y un buen análisis exploratorio de los datos. Emplea las estadísticas que consideres necesarias, así como cualquier

gráfico que permita aclarar y entender las conclusiones obtenidas. En el análisis exploratorio de los datos es importante que te emplees a fondo en aspectos estadísticos que sean significativos y aporten valor.

No olvides evaluar la capacidad de los modelos obtenidos, y decidir cuál emplearás finalmente para cada caso y porqué. Para finalizar, probarás el funcionamiento del sistema de inferencia inventando un cliente y prediciendo si se concedería el préstamo o no, y cuál sería su valor de riesgo asociado.

La entrega de esta actividad será únicamente el fichero .ipynb, que debe poder ser ejecutado secuencialmente sobre el fichero original de los datos, obteniendo los mismos resultados que estén en el archivo.