

Práctica 2: Limpieza y validación de los datos

Andrea González Vicario y M^a Dolores Higuera González

5 de enero de 2021

Contenidos

1. Descripción del dataset
2. Integración y selección de los datos
3. Limpieza de los datos.
4. Análisis de los datos.
5. Representación de los resultados a partir de tablas y gráficas.
6. Resolución del problema.

1. Descripción del dataset

Esta práctica tiene como objetivo el tratamiento de datos de un dataset con objetivo de aprender a identificar los datos relevantes para un proyecto analítico y usar las herramientas de integración, limpieza, validación y análisis.

Para ello, hemos elegido un dataset procedente de la web Kaggle (<https://www.kaggle.com/>) llamado “Credit Card customers”. Este dataset recoge los datos personales y de actividades bancarias de unos 10.000 clientes de un banco.

La intención de este estudio es conseguir crear un modelo con el que el gerente de banco pueda prever que clientes tienen más posibilidades de abandonar el banco. El conjunto de datos tiene un hándicap que dificulta mucho la obtención de un modelo eficaz, es por eso que el objetivo principal será conseguir los factores que tienen más peso en la pérdida de clientes.

Como datos personales recogidos de los clientes tenemos: edad, sexo, estado civil, estudios finalizados, salario...

Como información bancaria encontramos: tipo de tarjeta, antigüedad en el banco, meses de inactividad, transacciones, cantidad de las transacciones, límite de crédito...

2. Integración y selección de los datos

Se elimina la información de código interno de cliente y las dos últimas columnas
Continúa...