

Modelización de Problemas de Empresa

***Problema propuesto: Interpretabilidad en
Riesgo de Crédito***

Modelización de Problemas de Empresa – Problema Propuesto

Título: Interpretabilidad en Riesgo de Crédito

Descripción: Aplicaciones de machine learning (ML) y deep learning (DL) forman parte de nuestro día a día en numerosas aplicaciones y servicios. Sin embargo, hay algunas de estas aplicaciones que tienen un impacto directo sobre la vida de las personas o la economía de un país, un ejemplo claro son los modelos para la concesión de hipotecas, también llamados modelos de riesgo de crédito.

Por requisitos regulatorios, dichos modelos deben ser capaces de explicar a una persona por que le ha sido concedido o denegado el crédito solicitado. Motivado por ello, se plantea el siguiente problema:

La creación de un modelo interpretable con los datos proporcionados, donde la relación entre las variables input y el output sea simple y entendible por cualquier persona.

Detalles del desarrollo:

Para la resolución de este problema se dispone de un conjunto de datos etiquetados (train.csv) donde la variable a predecir es "loan_status". Dicha variable representa la probabilidad de impago de un cliente de una entidad financiera (donde el valor 1 representa impago).

Para ello, el participante deberá:

- Familiarizarse con aquellos modelos de ML interpretables (modelos lineales, arboles de decisión, etc).
- Familiarizarse con técnicas de preprocesamiento de datos utilizando técnicas de ML.
- Construir un pipeline de ML y entrenar el modelo con los datos proporcionados. La métrica para optimizar será el AUC.

Una vez desarrollado el modelo se deberá aplicar a los datos de testeo (test.csv) y preparar un archivo de entrega (solution_id.csv) con el formato indicado en "solution_random.csv".

Se valorarán positivamente la interpretabilidad del modelo y su *performance* sobre los datos de testeo. Sin embargo, la interpretabilidad será requisito indispensable para la correcta solución del problema y se espera una detallada explicación del modelo y preprocesamiento de los datos el día de la exposición.