Descripción proyecto

El proyecto trata de predecir si un prestatario de un banco entrará en una situación de default. Para empezar, se ha realizado un análisis exploratorio previo para ver que variables estaban más correlacionadas con la variable **not.fully.paid** (posibilidad de default) y se ha observado, entre otras, una correlación positiva del tipo de interés, la deuda sobre ingresos de los prestatarios y los requerimientos de nuevos préstamos en los 6 meses previos por parte de los deudores. Por su parte, observamos una correlación negativa del ratio FICO ( que mide la solvencia de un deudor) y el ratio otorgado por la compañía LendingClub.

Se han empleado tanto modelos de ML como RandomForest, XGBoost y LogisticRegression como redes neuronales para comprobar la accuracy del modelo. El resultado ha sido de 88.60%. Al evaluar la idoneidad del modelo con la ROC, el AOC ha salido un 90.1%, un resultado cercano a un clasificador perfecto que clasifica al 100% con lo cuál concluimos que se trata de un muy buen modelo. Además, a la hora de evaluar otras métricas hemos observado un recall de 98.5% que implica que los falsos negativos son ínfimos (0.62% concretamente), lo cual es también fundamental en el modelaje para un banco, pues minimiza el riesgo de que un individuo que según el modelo no entrará en default entré en verdad en default y ocasione pérdidas importantes a la entidad bancaria. La combinación de una buena accuracy y un gran recall para un banco maximiza la rentabilidad y le permite otorgar una buena rentabilidad al accionista.

José Antonio Suárez