

Proyecto de ML para reducir el riesgo crediticio

Introducció

El proyecto se basa en el análisis de los datos de los clientes de una institución bancaria portuguesa. En el banco los clientes pueden contraer préstamos, así como tomar depósitos a término. Este proyecto busca predecir si un cliente del banco cumplirá o no con el pago de sus préstamos lo cual es esencial para minimizar riesgos y mejorar la gestión segmentando a los clientes en función de su riesgo crediticio.

Objectius del Projecte

1. Quins són els objectius del negoci?

Se busca minimizar el incumplimiento de pagos de los préstamos por parte de los clientes, priorizando los créditos a clientes con bajo riesgo de incumplimiento crediticio.

2. Quines decisions o processos específics voleu millorar o automatitzar amb ML?

Al automatizar la identificación de clientes con historial de incumplimiento se evita la evaluación manual y subjetiva y se pueden proporcionar recomendaciones para políticas de crédito basadas en datos predictivos.

3. Es podria resoldre el problema de manera no automatitzada?

Sí, se podría a juicio del ejecutivo de cuenta, pero lo mismo puede introducir sesgos humanos y ser mas lento como poco justificable.

Metodologia Proposta

Para este problema necesitamos predecir si el cliente tiene un historial de incumplimiento crediticio (sí o no) respecto al pago de un préstamo (variable objetivo: default), lo cual es problema de clasificación binaria. En estos caso un **algoritmo de regresión logística** sería adecuado ya que es un modelo simple y facilmente interpretable. Sin embargo se tendrá que hacer un fuerte trabajo de pre-procesamiento de los datos para que el modelo funcione correctamente, como el realizado en la tarea 6.

Dado que el costo de incumplimiento crediticio ("default") para la institución bancaria es alto, debere evaluar el rendimiento del modelo y en base a la Matriz de Confusión priorizare las siguientes métricas:

*Precisión (Accuracy): Mide el porcentaje de predicciones correctas (verdaderos positivos y negativos) sobre el total. Como inicialmente mis clases estaban desbalanceadas me interesa sobremanera evaluar el modelo luego del balanceo que se realizo en el pre-procesamiento de datos.

* Sensibilidad (Recall): Proporción de casos positivos correctamente identificados. En este caso será la proporción de clientes que realmente incumplen (default = sí) identificados en el modelo lo cual es crucial para reducir los riesgos.

*F1-Score: Es una formulaque combina precisión y sensibilidad en una métrica única.

Por su lado también me interesara analizar los umbrales de probabilidad para asi ajustar la aceptación o rechazo de clientes en función de los riesgos de incumplimiento, a través de la curva ROC y AUC.

Finalmente será útil hacer una cross-validation para verificar que el modelo es robusto frente a distintas particiones de los datos.

Dades Disponibles

El banco cuenta con un conjunto de datos con información de 11.162 clientes y 17 variables, incluyendo datos demográficos (age, marital, education, job), información financiera (balance, default, deposit, housing, loan) y detalles de una campaña de marketing (day, duration, campaign, pdays, previous, contact, month, poutcome).

La variable objetivo / etiqueta será: "default" (valores: sí/no).

Las variables relevantes / características serán: 'balance', 'loan', 'pdays', 'previous' y 'deposit'.

Mètrica d'èxit del projecte

Reducir a menos del 1% los préstamos que entran en “default” anualmente.

Responsabilitats Ètiques i Socials

Como este proyecto implica el uso de datos personales y financieros de los clientes, es importante tener en cuenta la privacidad y protección de datos personales y cumplir con el RGPD, ser transparentes en cómo se usaran los datos, obtener el consentimiento explícito de los clientes, evitar sesgos o discriminación asegurando igualdad de oportunidades de acceso al crédito.