

12TA02_11 Ejercicio de Regresión y clasificación



Ejercicio Complejo: Análisis Predictivo y de Clasificación para la Rentabilidad de Clientes en una Empresa Financiera

Contexto:

Una empresa financiera quiere realizar un análisis exhaustivo de sus clientes para predecir tanto sus ingresos futuros como su probabilidad de rentabilidad. Para este propósito, han proporcionado un conjunto de datos con características clave de cada cliente. La empresa está interesada en implementar modelos avanzados que le permitan entender y optimizar su estrategia de fidelización.

Objetivo:

Realizar un análisis exploratorio y desarrollar modelos de predicción con el fin de:

1. **Predecir los ingresos futuros** de cada cliente utilizando un modelo de regresión lineal múltiple.
2. **Clasificar si un cliente será rentable o no** usando un modelo de regresión logística y otro de clasificación (árbol de decisión o bosque aleatorio).
3. **Evaluar y comparar el rendimiento de los modelos** en ambos objetivos.

Enunciado

1. **Análisis Exploratorio de Datos (EDA):**
 - Realiza un análisis de correlación para evaluar la relación entre las variables. ¿Cuáles parecen estar más relacionadas con los ingresos futuros y con la rentabilidad?
 - Visualiza las distribuciones de `edad`, `ingreso_anual`, `saldo_promedio`, `gastos_mensuales` y `historial_crediticio`.
 - Crea gráficos de dispersión para explorar las relaciones entre `ingreso_anual`, `saldo_promedio` y `ingresos_futuros`.
2. **Preprocesamiento de Datos:**
 - Normaliza las variables `edad`, `ingreso_anual`, `saldo_promedio`, `gastos_mensuales` y `historial_crediticio` para su uso en los modelos.
 - Divide los datos en conjuntos de entrenamiento (80%) y prueba (20%).
3. **Regresión Lineal Múltiple:**
 - Desarrolla un modelo de regresión lineal múltiple para predecir `ingresos_futuros` utilizando las variables `edad`, `ingreso_anual`, `saldo_promedio`, `gastos_mensuales` y `historial_crediticio`.
 - Evalúa el modelo utilizando el error cuadrático medio (MSE) y el coeficiente de determinación (R^2) en el conjunto de prueba.
 - Analiza los coeficientes de regresión para entender el impacto de cada variable en los ingresos futuros.

12TA02_11 Ejercicio de Regresión y clasificación



4. Clasificación de Rentabilidad con Regresión Logística:

- Usa las mismas variables (`edad`, `ingreso_anual`, `saldo_promedio`, `gastos_mensuales`, `historial_crediticio`) para construir un modelo de regresión logística que clasifique si un cliente será rentable.
- Evalúa el rendimiento del modelo en el conjunto de prueba usando métricas como precisión, recall, y la matriz de confusión.
- Interpreta los coeficientes del modelo y analiza cómo cada variable afecta la probabilidad de rentabilidad.

5. Clasificación de Rentabilidad con Árbol de Decisión / Bosque Aleatorio:

- Desarrolla un modelo de clasificación de rentabilidad alternativo usando un árbol de decisión o un bosque aleatorio.
- Ajusta el hiperparámetro `max_depth` para el árbol de decisión y `n_estimators` para el bosque aleatorio (si lo usas) utilizando validación cruzada para mejorar el rendimiento.
- Evalúa el modelo en el conjunto de prueba y compara las métricas de precisión, recall y F1 con las obtenidas en la regresión logística.

6. Comparación y Conclusiones:

- Compara el rendimiento de los modelos de clasificación (regresión logística y árbol de decisión/bosque aleatorio). ¿Cuál muestra un mejor rendimiento en términos de precisión y F1?
 - Compara el modelo de regresión lineal múltiple y el de clasificación en términos de interpretabilidad y utilidad para la empresa.
 - Redacta una conclusión general indicando qué variables y modelos son los más útiles para optimizar la estrategia de rentabilidad de los clientes.
-