**TRABAJO DE APLICACIÓN**

CASO: CRÉDITO BANCARIO

En el ámbito bancario, la concesión de créditos es una actividad central que implica evaluar el riesgo crediticio asociado con cada solicitante. Los bancos recopilan una gran cantidad de datos sobre los clientes potenciales, que incluyen información financiera, historial de crédito, ingresos, deudas, y otros factores relevantes. Utilizando estos datos, los bancos desarrollan modelos de regresión para predecir la probabilidad de incumplimiento de los préstamos, lo que les ayuda a tomar decisiones informadas sobre la concesión de créditos.

Problemática: A pesar de los avances en la modelización de riesgos crediticios, los bancos todavía enfrentan desafíos para identificar de manera precisa las variables más influyentes y desarrollar los modelos de regresión más efectivos. Algunos de estos desafíos incluyen:

1. Identificación de variables relevantes: Determinar qué variables tienen el mayor impacto en el riesgo crediticio puede ser complicado, especialmente cuando se consideran múltiples factores interrelacionados. La falta de una comprensión clara de qué variables son más predictivas puede afectar la precisión del modelo.
2. Elección del mejor modelo: Determinar el mejor modelo puede resultar un problema si no se seleccionada adecuadamente las variables e indicadores
3. Interpretación del modelo: Los modelos de regresión pueden ser inherentemente difíciles de interpretar, lo que dificulta la comprensión de cómo cada variable contribuye al riesgo crediticio. Una comprensión incompleta de la relación entre las variables y el riesgo puede limitar la capacidad de los bancos para tomar decisiones fundamentadas.

Objetivos: Para abordar estas problemáticas y mejorar la eficacia de los modelos de regresión de riesgo crediticio, se plantean los siguientes objetivos:

1. Identificar las variables más relevantes: Utilizar técnicas de análisis exploratorio de datos y selección de características para identificar las variables que tienen el mayor impacto en el riesgo crediticio.

Con un modelo de regresión puedes identificar qué variables son significativas y que variables deben ser excluidas.

1. Desarrollar modelos de regresión robustos: Explorar diferentes algoritmos de regresión y técnicas de regularización para desarrollar modelos que sean precisos y generalizables.

Comparar modelos, primer modelo con el segundo modelo.

1. Interpretar el modelo: Emplear métodos de interpretación de modelos, como gráficos de importancia de características y análisis de contribución de variables, para comprender mejor cómo cada variable influye en el riesgo crediticio.

Interpretar y evaluar modelo.

1. Evaluar el rendimiento del modelo: Utilizar métricas de evaluación del rendimiento, como la precisión, la sensibilidad y la especificidad, para evaluar y comparar la eficacia de los modelos desarrollados.

Al lograr estos objetivos, los bancos pueden mejorar su capacidad para evaluar el riesgo crediticio de manera precisa y tomar decisiones informadas sobre la concesión de créditos, lo que les ayuda a gestionar mejor su cartera de préstamos y minimizar las pérdidas financieras.

Ante estas expectativas se pide a usted utilizar la data crédito.xlsx y desarrollar dos modelos: uno de regresión y otro de clasificación (regresión logística).

Las variables se encuentran a continuación:

|  |  |
| --- | --- |
| **Monto (Y1)** | Monto de crédito |
| **Credito (Y2)** | Si se le debe dar el crédito al cliente (Si, No) |
| **Sexo** | Sexo del cliente (Masculino, femenino) |
| **E.Civil** | Estado civil del cliente |
| **Educacion** | Grando de instrucción |
| **Prioridad** | Si el cliente es prioritario (Si, No) |
| **Edad** | Edad del cliente |
| **Tarjetas** | Cantidad de tarjetas de crédito del cliente |
| **Deuda** | Deuda del cliente |
| **Saldo** | Saldo del cliente |
| **CrediScore** | Puntaje de crédito del cliente |
| **Años\_empleo** | Años de empleo del cliente |
| **Ingresos** | Ingresos del cliente |

Variable **Monto (Y1):** sirve para regresión lineal múltiple.

Variable **Credito (Y2):** para regresión logística.

**Variable X:**

El resto de las 11 variables X, son predictoras. Usan para ambos modelos.