|  |  |
| --- | --- |
|  | **Maestría en Ingeniería de Información (MINE)**  **Ingeniería de Sistemas y Computación**  **Curso: MINE 4101-Ciencia de datos aplicada**  **Taller combinación de técnicas** |

**Caso: financiera Alpes**

Financiera de los Alpes, es una empresa del sector financiero, que ofrece tres productos: TDC (tarjeta de crédito), CC (crédito de consumo o de libre inversión) y SBG (sobregiro). Esa empresa, desea generar estrategias comerciales que permitan reducir el número de clientes perdidos por mal comportamiento crediticio. En particular, se entiende como mal comportamiento, clientes con una altura de mora mayor a 90 días en algunos de sus productos. Cuando esta situación ocurre, la entidad bloquea todos los productos de ese cliente y no es posible cambiar su estado, llevando a la pérdida definitiva del cliente. En la actualidad, financiera de los Alpes presenta un porcentaje relativamente alto de clientes perdidos por su mal comportamiento crediticio y desea generar estrategias que permitan mejorar esta situación.

Con el fin de abordar esa problemática, Financiera de los Alpes, ha puesto a su disposición una **muestra de datos** (archivo adjunto en SICUA) que permite identificar las características de los clientes que pueden ser más propensos a perderse por mal comportamiento crediticio.

De acuerdo con la descripción del caso, responda las siguientes preguntas:

1. Describa el resultado del perfilamiento de datos e indique para cada variable si es categórica o no.
2. Describa e ilustre los problemas de calidad de los datos que pueden afectar los análisis posteriores, justifique su respuesta.
3. Realice los ajustes que considere necesarios sobre los datos y justifique sus decisiones.
4. Describa el entendimiento que tiene del negocio a partir de los datos suministrados.
5. Plantee 3 hipótesis de interés para el negocio que al resolverlas generen un beneficio. Estas hipótesis deben considerar múltiples variables.
6. Seleccione una hipótesis del punto anterior, escoja y justifique la técnica o modelo a utilizar para validarla (por ejemplo, regresión lineal, , árboles de decisión, etc.). Cualquiera que sea su elección, justifíquela basado en los criterios que considere relevantes para su caso en particular. Por ejemplo, puede utilizar criterios de disponibilidad en su software estadístico, conocimiento personal, explicabilidad del modelo, complejidad de implementación, etc.
7. Describa el proceso general para la construcción de su modelo, en particular, utilice los siguientes puntos como guía:
   1. ¿Qué variables decidió incluir y por qué?
   2. ¿Realizó transformaciones sobre los datos? Por ejemplo, algunos algoritmos requieren transformar variables categóricas a variables binarias (una por cada categoría), o decidió convertir una variable continua a categórica (como transformar edad en años a “joven” vs. “adulto”)
   3. En ejercicios de pronóstico, es importante dividir el conjunto de datos en “entrenamiento” y “test”. Explique el por qué de esta separación y cómo la implementa en su ejercicio. Sea explícito sobre qué datos incluyó para cada uno (fechas, cantidad de registros, etc.)
   4. Si incluye ajustes de hiperparámetros en su modelo para mejorar el pronóstico, describa el proceso. Si no lo hace, explique la intuición de hacerlo y si su modelo tiene la posibilidad de ser ajustado de esta manera
8. Muestre los resultados predictivos de su modelo. ¿Qué métricas de desempeño (R cuadrado, recall, precisión, matriz de confusión, etc) considera relevantes para su modelo? Explique la intuición y el por qué de la utilización de esta métrica, mostrando a través de gráficas, tablas o cualquier herramienta que considere necesaria, los resultados más relevantes. ¿Considera “buenos” los resultados del modelo?
9. ¿De qué forma se podría utilizar en el negocio los resultados del ejercicio? Sea explícito en la forma en que esto se haría, con qué objetivos concretos y cómo se tomarían las decisiones

**Descripción de variables**

|  |  |
| --- | --- |
| **Variable** | **Descripción** |
| CUENTA\_TDC | Número de cuenta TDC |
| ALTURA\_MORA\_TDC | Días de mora actual en la TDC |
| HABITO\_PAGO\_TDC | Hábito de pago en la TDC |
| Edad | Edad en años del cliente |
| ALTURA\_MORA\_CC | Días de mora actual en el crédito fijo |
| ALTURA\_MORA\_SBG | Días de mora en el sobregiro |
| ALTURA\_MOR\_MAX | MAX (ALTURA\_MORA\_TDC; ALTURA\_MORA\_CC; ALTURA\_MORA\_SBG), son los días de mora más altos que tiene el cliente. |
| SALDO\_TDC | Saldo actual de la TDC (lo que el cliente debe en su tarjeta de crédito) |
| SALDO\_CC | Saldo actual del crédito fijo (lo que el cliente debe en su crédito de consumo) |
| SALDO\_SBG | Saldo actual del sobregiro (lo que el cliente debe en su sobregiro) |
| CUPO\_TDC | Cupo actual tarjeta de crédito |
| CUPO\_CC | Cupo actual crédito de consumo |
| CUPO\_SBG | Cupo actual sobregiro |
| SALDO\_TOTAL | SUMA (SALDO\_TDC SALDO\_CC SALDO SBG), es la deuda total que tiene el cliente con la entidad financiera |
| Y | 1: Si el cliente se pierde por mal comportamiento y 0 en caso contrario |
| UTIL\_TDC | Porcentaje de uso en la TDC, se define como SALDO TDC/CUPO TDC |
| CUPO\_TOTAL | SUMA (CUPO\_TDC; CUPO\_CC; CUPO\_SBG), es el cupo total que tiene el cliente en todos los productos. |
| UTIL\_TOTAL | Porcentaje de uso de todos los productos del cliente, se define como SALDO\_TOTAL/CUPO\_TOTAL |
| SUMA\_ALTURA\_MORA | SUMA (ALTURA\_MORA\_TDC; ALTURA\_MORA\_SBG; ALTURA\_MORA\_CC), es la suma de las alturas de mora de todos los productos financieros |
| MORA\_MAX\_N\_12 | Altura de mora máxima que ha tenido el cliente en los últimos 12 meses (sin incluir el mes actual) |
| MAX\_CUPO\_TDC\_3 | Máximo cupo que ha tenido el cliente en su TDC en los últimos 3 meses |
| MAX\_CUPO\_TDC\_6 | Máximo cupo que ha tenido el cliente en su TDC en los últimos 6 meses |
| MAX\_CUPO\_TDC\_12 | Máximo cupo que ha tenido el cliente en su TDC en los últimos 12 meses |
| NUM\_MESES\_SIN\_MORA\_3 | Número de meses sin mora en los últimos 3 meses (sin incluir el mes actual) |
| NUM\_MESES\_SIN\_MORA\_6 | Número de meses sin mora en los últimos 6 meses (sin incluir el mes actual) |
| NUM\_MESES\_SIN\_MORA\_12 | Número de meses sin mora en los últimos 12 meses (sin incluir el mes actual) |
| estrato | Estrato socioeconómico |
| NOMBRE\_CIUDAD | Ciudad donde el cliente sacó su primer producto |
| DEPARTAMENTO | Departamento donde el cliente sacó su primer producto |
| ACT\_ECO | Actividad económica del cliente |
| SEXO | Género del cliente |
| NIVEL\_EST | Nivel de estudios del cliente |