

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

MAESTRÍA EN CIENCIAS DE DATOS

APRENDIZAJE AUTOMÁTICO

MAESTRO: JOSÉ ALBERTO BENAVIDES VÁZQUEZ

TAREA # 1

ALUMNO: EDWIN MARTÍN ROMERO SILVA

MATRÍCULA: 1731276

Tarea #1

Definir un conjunto de datos

La base de datos que escogí se llama **BD_solicitudes**, contiene todas las solicitudes de crédito automotriz del Banco en el que trabajo, desde enero 2018 hasta julio 2023. Además de contener características de los clientes como edad, estado civil, antigüedad siendo cliente del banco etc., contiene variables del historial crediticio de los clientes.

Todas las variables fueron calculadas/recolectadas al momento de la solicitud de crédito.

Esta base no la escogí con fines de cruzar bases de datos como la de la materia anterior. Esta tiene el propósito de ser utilizada para entrenar modelos que permitan predecir.

La variable más interesante para predecir en esta base es la variable de impago, es decir buscaremos predecir si el cliente pagará o no un crédito. Las variables contenidas en esta base son:

- **Folio:** ID de la solicitud
- **Fecha_solicitud:** Fecha de solicitud
- **Default60DD12M:** Bandera que indica si el cliente llegó a 60 días de atraso en los primeros 12 meses desde que se origina el crédito.
- **Var_hit:** Bandera que indica si el cliente cuenta con historial crediticio al momento de la solicitud.
- **KO:** Bandera que indica si el cliente cuenta con más de 90 días de atraso en algún crédito vigente.
- **Pct_uso_rev:** Porcentaje de utilización créditos revolventes
- **Antig_Bca_buro:** Antigüedad en buró de crédito en Bancos
- **Meses_cierre_tdc:** Meses desde el cierre más reciente de TDC
- **Num_consultas_6m:** Número de consultas a buró de crédito en los últimos 6 meses
- **Estado civil:** Estado civil del solicitante
- **Estado laboral:** Estado donde trabaja el cliente
- **Pct_cred_corriente:** Porcentaje de créditos al corriente

Justificación de la elección

Mi trabajo consiste en hacer modelos para aceptar/rechazar solicitudes de crédito, estos modelos tienen el objetivo de predecir el impago. Y ya que en esta materia entrenaremos algunos modelos, creo que me servirá mucho utilizar una base de datos de solicitudes de crédito, incluyendo una variable que indique su performance (Default60DD12M).

Objetivos

Al igual que en otras materias, mi objetivo es aprender, en este caso me gustaría mejorar la forma en que entreno los modelos y conocer nuevos algoritmos de ML.

Nota: Pedí permiso en el trabajo para utilizar esta información, en caso de que me nieguen el permiso, tendré que cambiar la base de datos, pero utilizaría una de Kaggle con el mismo enfoque.