

Trabajo Práctico Final

Modelo de Predicción de Riesgo Crediticio

Sistemas de Soporte para la Toma de Decisiones

2024 - UNPSJB





Introducción

Los modelos estadísticos pueden ayudar a las entidades financieras para la toma de decisiones a la hora de otorgar un préstamo a un cliente potencial. Estos modelos intentan predecir el comportamiento futuro de un potencial cliente en base a su comportamiento histórico. Dependiendo del modelo, puede ser el caso que genere un puntaje de crédito dentro de un rango predefinido o que clasifiquen al cliente dentro de una categoría preestablecida.

En el siguiente trabajo se presenta un dataset de unos 32.000 registros históricos que simula un conjunto de datos de clientes obtenidos de bureaus financieros. Cada muestra representa un préstamo otorgado a un cliente. Por cada préstamo se cuenta con un conjunto de características del cliente y del préstamo y una columna que indica el estado actual del préstamo otorgado. Este estado puede ser '0', significa que está todo correcto, o '1', significa que el préstamo está en mora.

Nombre de la columna	Descripción
person_age	Edad de la persona
person_income	Ingreso anual en USD
person_home_ownership	Indica si es propietario, propietario con hipoteca o alquiler
person_emp_lenght	Años de antigüedad en el empleo actual
loan_intent	Objetivo del préstamo
loan_grade	Grado del préstamo
loan_amnt	Monto del préstamo
loan_int_rate	Tasa de interés
loan_status	Estado del préstamo, donde 0 es préstamo pagado y 1 es préstamo en mora: Columna resultado
loan_percent_income	Porcentaje que representa el préstamo con respecto al ingreso anual
cb_person_default_on_file	Indica si la persona ha estado en mora anteriormente
cb_person_cred_hist_lenght	Años de historial crediticio

El objetivo del trabajo es desarrollar un modelo que permita predecir si un cliente potencial que tome un préstamo tiene altas probabilidades de caer en mora. Es decir, que intente predecir si el cliente va a poder pagar el préstamo en tiempo y forma.

Un modelo con estas características permitiría a un oficial de cuentas de una entidad financiera decidir de forma asistida si conviene o no aceptar al cliente potencial en la entidad.



Actividades

1. Análisis del dataset provisto.
 - a. Explore los datos para detectar outliers y datos nulos.
 - b. Utilice gráficos para entender el comportamiento de los datos.
 - c. ¿Es posible identificar características que influyen más que otras en la decisión?
 - d. ¿Es posible identificar relación entre algunas de las características?
 - e. ¿El dataset se encuentra balanceado?
2. Pre procesamiento de los datos
 - a. Realice el pre procesamiento de datos necesario para desarrollar los modelos indicados en el punto siguiente.
 - b. ¿Es aconsejable en este caso utilizar alguna técnica de reducción de dimensiones?
3. Desarrolle los siguientes modelos y compare los resultados.
 - a. Un árbol de decisión, un método de ensamble y una red neuronal artificial.
 - b. Otro modelo a elección.
 - c. ¿Qué métricas son las más adecuadas para medir cada modelo?
 - d. Grafique los resultados.
4. Seleccione el modelo con los mejores resultados y desarrolle una aplicación web donde se pueda simular un análisis de crédito sobre un cliente potencial. La aplicación debe ser para uso de un usuario final utilizando los criterios de una buena visualización.
5. Realice un informe escrito y una presentación para la explicación final del trabajo en clase.
 - El informe debe estar escrito con el siguiente formato:
 - i. Realizar una portada con logo de la Universidad, título del informe, nombre de la cátedra, nombre del alumno, año y responsables de cátedra.
 - ii. Todo el texto debe estar justificado
 - iii. Utilizar títulos para destacar el multinivel de lo escrito (formato de Word o la herramienta que se utilice)
 - iv. Para los párrafos utilizar letra 'Arial' con tamaño 12.
 - v. Utilizar encabezado y pie de página numerando todas las páginas.
 - vi. Realizar una revisión final de ortografía y gramática.