

Proyecto Entrega 1

Default of Credit Card Clients Dataset

Nilson Fernando Suarez Hernández



1 8 0 3

Introducción a la Inteligencia Artificial

Facultad de Ingeniería

Universidad de Antioquia

Julio de 2022/1

1) Problema predictivo a resolver

La resolución del problema tendrá como objetivo predecir si la persona (cliente) va a pagar o no va a pagar al mes siguiente de la fecha de análisis, cuya variable principal sería 'default.payment.next.month: Default payment (1=yes, 0=no)'.

2) Dataset a utilizar

UCI_Credit_Card.csv

El conjunto de datos contiene información sobre impagos, factores demográficos, datos crediticios, historial de pagos y extractos de facturas de clientes de tarjetas de crédito en Taiwán desde abril de 2005 hasta septiembre de 2005.

3) Métricas de desempeño requeridas (de machine learning y de negocios)

A continuación, se presentarán un listado de métricas que serán posibles candidatos para utilizar, de las cuales se filtrará una o dos la cual será la técnica final a utilizar.

De machine learning

- Regresión logística: Se para predecir el resultado de una variable categórica en función de las variables independientes o predictoras.
- Random Forest: Construye árboles de decisión sobre diferentes muestras y toma su voto mayoritario para la clasificación y la media en caso de regresión.
- Árbol de decisión: Puede realizar tareas de clasificación o regresión
- K Nearest Neighbors: Hace clasificaciones o predicciones sobre la agrupación de un punto de datos individual.
- Support vector machine: Es capaz de predecir si un punto nuevo (cuya categoría desconocemos) pertenece a una categoría o a la otra.
- Regresión lineal: Calcula la puntuación de error para resumir la habilidad predictiva de un modelo
- MAE (Mean Absolute Error): Mide la precisión de la diferencia entre dos variables continuas.
- RMSE: Para medir la precisión de las variables continuas
- Precisión_score: La precisión es la relación $tp / (tp + fp)$ donde tp es el número de verdaderos positivos y fp el número de falsos positivos.

De negocios

- Brier Score: mide la diferencia media al cuadrado entre la probabilidad predicha y el resultado real.
- KMV: Calcula la probabilidad de incumplimiento basado en la información obtenida
- Z score: Calcula la probabilidad de que incumpla el pago
- ADD: Mide los días de mora de un cliente

Se aspira a alcanzar un mínimo de 70% en las pruebas, siendo evaluado con la métrica seleccionada.

4) Un primer criterio de cuál sería el desempeño deseable en producción

Se pretende alcanzar un acierto de por lo menos un 70%, de esta manera garantizar un buen desempeño del modelo, y por consiguiente una buena predicción que ayude a una buena toma de decisión.

REFERENCIAS

LEARNING, U. M. (2016). *Default of Credit Card Clients Dataset*. Kaggle.
https://www.kaggle.com/datasets/uciml/default-of-credit-card-clients-dataset?select=UCI_Credit_Card.csv