

INTRODUCCIÓN A LA INTELIGENCIA ARTIFICIAL

PROYECTO IA

JOSE CARLOS ORTIZ PADILLA

1003059949

Ingeniería de Sistemas

ALEXANDER VALENCIA DELGADO

71372112

Ingeniería de sistemas

JORGE ERNESTO VILLAMIZAR CABEZA

091487510

Ingeniería Industrial

UNIVERSIDAD DE ANTIOQUIA

MEDELLIN

2023

Home Credit Default Risk

Enlace de la competición en kaggle

<https://www.kaggle.com/c/home-credit-default-risk/overview>

1. Describir el problema

Muchas personas tienen dificultades para obtener préstamos debido a antecedentes crediticios insuficientes o inexistentes. Y, desafortunadamente, los prestamistas poco confiables a menudo se aprovechan de esta población.

El problema que se aborda en esta competencia es la predicción del riesgo de impago de préstamos concedidos por Home Credit a solicitantes que tienen dificultades para acceder a préstamos bancarios.

El riesgo de impago es una preocupación importante para los prestamistas, ya que puede tener consecuencias financieras significativas. Home Credit busca mejorar su capacidad para evaluar el riesgo de impago y tomar decisiones informadas sobre la aprobación o denegación de solicitudes de préstamo. Los modelos de aprendizaje automático desarrollados en esta competencia podrían ayudar a la empresa a tomar decisiones más precisas y mejor informadas, lo que podría reducir su riesgo de pérdidas financieras y mejorar su rentabilidad.

2. Dataset a utilizar

vamos a usar el dataset de kaggle esta competición en este caso hacemos referencia a la tabla `application_test`

(https://www.kaggle.com/competitions/home-credit-default-risk/data?select=installments_payments.csv), que tiene 48.745 número de muestras o instancias (cliente que ha solicitado préstamo) y tales columnas, como es un listado muy grande que corresponde a 121 columnas de las cuales hemos elegido 30 de suma importancia, (AMT_ANNUITY, AMT_CREDIT, AMT_GOODS_PRICE, APARTMENTS_AVG, BASEMENTAREA_AVG, CODE_GENDER, DAYS_BIRTH, DAYS_EMPLOYED, DAYS_ID_PUBLISH, DAYS_LAST_PHONE_CHANGE, DEF_30_CNT_SOCIAL_CIRCLE, DEF_60_CNT_SOCIAL_CIRCLE, EXT_SOURCE_1, EXT_SOURCE_2, EXT_SOURCE_3, FLAG_DOCUMENT_3, FLAG_DOCUMENT_8, FLAG_DOCUMENT_13, FLAG_DOCUMENT_16, FLAG_DOCUMENT_18, FLAG_OWN_CAR, FLAG_OWN_REALTY, NAME_CONTRACT_TYPE, NAME_EDUCATION_TYPE, NAME_FAMILY_STATUS, NAME_HOUSING_TYPE, NAME_INCOME_TYPE, OCCUPATION_TYPE, REGION_POPULATION_RELATIVE, TOTALAREA_MODE).

3. Métricas de desempeño

Las métricas de aprendizaje automático que se utilizan para evaluar el rendimiento de los modelos son la AUC-ROC, La curva AUC-ROC (área bajo la curva característica de operación del receptor) mide la capacidad del modelo para distinguir entre las clases positivas y negativas. Es una métrica útil para problemas de clasificación binaria, donde la variable objetivo es categórica y tiene dos posibles valores

4. Criterio de desempeño deseado

Si el modelo se utiliza para tomar decisiones de préstamo en tiempo real, la velocidad de predicción y la capacidad del modelo para mantener una baja tasa de falsos positivos y falsos negativos pueden ser factores críticos para su desempeño. Si se utiliza para análisis de riesgo a largo plazo, la capacidad del modelo para predecir con precisión el riesgo de impago a lo largo del tiempo puede ser más importante que su capacidad para distinguir entre las clases positivas y negativas.

En general, el desempeño deseable de un modelo en producción se medirá en función de su capacidad para mejorar la precisión y la eficiencia de los procesos de negocio de la empresa, lo que puede requerir el equilibrio de varias métricas y objetivos comerciales clave.