

INTRODUCCIÓN A LA INTELIGENCIA ARTIFICIAL

PROYECTO IA

ALVARO JAVIER TALAIGUA ROSARIO
1196969101
INGENIERÍA MECÁNICA

DUBAN VILLA BEDOYA
71225370
INGENIERÍA MECÁNICA

WILSON ARBEY CHAVERRA MEJIA
8064138
INGENIERÍA MECÁNICA

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 desafío que se plantea en esta competencia se centra en predecir la probabilidad de que los préstamos otorgados por Home Credit a individuos con dificultades para acceder a préstamos bancarios no sean reembolsados. Esta incertidumbre sobre el cumplimiento de los pagos representa una inquietud importante para los prestamistas, ya que puede tener repercusiones financieras considerables. Home Credit está buscando fortalecer su capacidad para evaluar este riesgo y tomar decisiones más fundamentadas sobre si aprobar o rechazar las solicitudes de préstamo. Los modelos de aprendizaje automático desarrollados en esta competencia tienen el potencial de asistir a la empresa en la toma de decisiones más precisas y fundamentadas, lo que podría disminuir su exposición al riesgo de pérdidas económicas y, al mismo tiempo, 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 empleadas en el aprendizaje automático para evaluar el desempeño de los modelos incluyen la AUC-ROC. La AUC-ROC, que significa "área bajo la curva de la característica operativa del receptor", se utiliza para medir la habilidad del modelo en diferenciar entre las clases positivas y negativas. Esta métrica resulta especialmente valiosa en situaciones de clasificación binaria, donde la variable objetivo es categórica y presenta únicamente dos posibles valores.

4. Criterio de desempeño deseado

Si empleamos el modelo en tiempo real para tomar decisiones de préstamo, la rapidez en la generación de predicciones y la habilidad del modelo para minimizar los errores de falsos positivos y falsos negativos se vuelven esenciales para su efectividad. No obstante, si se utiliza para realizar análisis de riesgo a largo plazo, su capacidad para prever con exactitud el riesgo de incumplimiento en el transcurso del tiempo podría tener una relevancia superior a su destreza para distinguir entre las clases positivas y negativas.

En términos generales, el rendimiento ideal del modelo en un entorno de producción se evaluará considerando su aptitud para mejorar la precisión y la eficacia de los procesos empresariales. Esto podría implicar la necesidad de equilibrar múltiples métricas y objetivos comerciales fundamentales.