Deploying a Credit Risk Application Capstone Project

Build an algorithm for credit scoring

# Description

Los bancos y empresas financieras, a la hora de efectuar una operación crediticia con un determinado cliente, se enfrentan al problema de la incertidumbre. Simplemente ellos no saben con plena certeza, si el cliente va cumplir o no con sus pagos, a esto se le suele llamar riesgo crediticio. Es decir, el riesgo de crédito es la posible pérdida que asume el banco o entidad financiera como consecuencia del incumplimiento de las obligaciones contractuales que incumben a las contrapartes con las que se relaciona.

Por ello, comúnmente los bancos diseñan diferentes formas o estrategias (reglas de negocio o filtros) para clasificar/discriminar a sus clientes, y así, disminuir el riesgo de la cartera de crédito, al intentar colocar sus recursos (fondos prestables) en operaciones crediticias de bajo riesgo, o evitar colocarlos en operaciones de alto riesgo.

El objetivo del presente proyecto es brindar y documentar un ejemplo de cómo se puede abordar el problema descrito anteriormente, con un algoritmo de aprendizaje automático. De forma que el algoritmo aprenda de un conjunto de datos de entrenamiento las características de los clientes que no pagan sus préstamos, y, de acuerdo a lo aprendido, brinde una nota de crédito, que pueda permitir a la entidad discriminar de manera efectiva entre clientes “buenos” y clientes “malos”, y así, al integrar este algoritmo en su proceso de aprobación, lograr disminuir los costos por riesgo crediticio. Adicionalmente, se mostrará cómo usar Amazon Sagemaker, AWS Lambda y API Gateway, para crear REST API, con la cual se puede establecer comunicación con el modelo, y así recibir solicitudes y enviar respuestas en tiempo real.

# Setup

Python 3.6

Inbalanced-Learn, NumPy, Pandas, Scikit-Learn

# Data

The data will be obtained from the following Kaggle source:

* <https://www.kaggle.com/ranadeep/credit-risk-dataset>

It contains 887,379 examples and 80 columns.

# Tasks

Para completar este proyecto será necesario completar las siguientes tareas.

1. Definir una variable objetivo
2. Realizar un análisis exploratorio de los datos, para así, además de comprender el conjunto de datos en general, determinar cuáles columnas del conjunto de datos pueden ser útiles para crear el algoritmo.
3. Seleccionar el algoritmo que mejor se ajuste a los datos de validación
4. Reportar el rendimiento del algoritmo seleccionado usando los datos de prueba.
5. A partir de las estimaciones de probabilidad del modelo seleccionado, crear una escala o nota de crédito que pueda ser interpretada por sus usuarios.
6. Poner en funcionamiento el modelo escogido en un endpoint de Amazon Sagemaker.
7. Crear una función en AWS lambda, la cual pre procesará los datos que se enviaran al endpoint
8. Crear un API en API Gateway de Amazon, que se podrá utilizar para establecer comunicación con el modelo.

## Blog Post in Medium

Se documentara todo el proceso de creación de la aplicación en mi cuenta personal de Medium.

## Test and measure performance

Se utilizaran diferentes métricas para evaluar el algoritmo, en una primera fase de selección del algoritmo, se utilizará principalmente el Kolmogorov-Statistic, y el “ROC-AUC score”. Posteriormente se utilizarán criterios para determinar el rendimiento del algoritmo agrupando los clientes por diferentes de niveles de riesgo.