



UNIVERSIDAD AUTÓNOMA DE OCCIDENTE



SOFTWARE MODELADOR DE IMPAGOS SMI

Angie Tatiana Aparicio
Jhonn Sebastian Robles
Jhon Brandon Idarraga
Harold Muñoz

Esp. Analítica en Big Data
UAO

Septiembre, 2024



PARTE 1: EQUIPO Y MARCO DE TRABAJO DEL PROYECTO

1. Diccionario de datos
2. Arquitectura de la Solución (Diagrama final)
3. Código Fuente
4. Interfaz Gráfica de Usuario
5. Video demostración de la solución desarrollada (video-captura de pantalla)
6. Presentación del Entregable (10 min) (demo en vivo)



1. Diccionario de datos

El diccionario de datos proporciona una descripción detallada de los campos utilizados en el proyecto, ayudando a garantizar una comprensión clara y precisa de los datos manejados a lo largo de todo el proceso.

Este recurso es fundamental para asegurar la consistencia y coherencia en el uso de los datos facilitando su análisis. A continuación, se presenta el diccionario de datos, el cual será una herramienta clave para el equipo y cualquier usuario que trabaje con el proyecto.

URL del Data Set:

<https://www.kaggle.com/datasets/uciml/default-of-credit-card-clients-dataset>

Nombre:

Default of Credit Card Clients Dataset.

Descripción de los campos:

SALIDA:

- **default.payment.next.month:** Sigue con la deuda a final de mes (1=Si, 0=No)

ENTRADA:

- **ID:** ID de cada cliente
- **LIMIT_BAL:** cantidad del crédito otorgado en dólares (incluye crédito individual y familiar / complementario)
- **SEX:** género (1=hombre, 2=mujer)
- **EDUCATION:** (1 = posgrado, 2 = universidad, 3 = secundaria, 4 = otros, 5 = desconocido, 6 = desconocido)
- **MARRIAGE:** estado civil (1=casado, 2=soltero, 3=otro)
- **AGE:** edad en años
- **PAY_0:** Estado de reembolso en Septiembre de 2005 (-1 = pagó debidamente, 1 = retrasó el pago durante un mes, 2 = retrasó en el pago durante dos meses, ... 8 = retraso en el pago durante ocho meses, 9 = retraso en el pago durante nueve meses y más)
- **PAY_2:** Estado de reembolso en agosto de 2005 (escala igual a la anterior)
- **PAY_3:** Estado de reembolso en julio de 2005 (escala igual a la anterior)
- **PAY_4:** Estado de reembolso en junio de 2005 (escala igual a la anterior)
- **PAY_5:** Estado de reembolso en mayo de 2005 (escala igual a la anterior)
- **PAY_6:** Estado de reembolso en abril de 2005 (escala igual a la anterior)



Figure 1: Descripción de los datos

• ENTRADAS (CONTINUACIÓN):

- **BILL_AMT1:** cantidad del extracto de la factura en septiembre de 2005 (dólar NT)
- **BILL_AMT2:** cantidad del extracto de la factura en agosto de 2005 (dólar NT)
- **BILL_AMT3:** cantidad del extracto de la factura en julio de 2005 (dólar NT)
- **BILL_AMT4:** cantidad del extracto de la factura en junio de 2005 (dólar NT)
- **BILL_AMT5:** cantidad del extracto de la factura en mayo de 2005 (dólar NT)
- **BILL_AMT6:** cantidad del extracto de la factura en abril de 2005 (dólar NT)
- **PAY_AMT1:** cantidad del pago anterior en septiembre de 2005 (dólar NT)
- **PAY_AMT2:** cantidad del pago anterior en agosto de 2005 (dólar NT)
- **PAY_AMT3:** cantidad del pago anterior en julio de 2005 (dólar NT)
- **PAY_AMT4:** cantidad del pago anterior en junio de 2005 (dólar NT)
- **PAY_AMT5:** cantidad del pago anterior en mayo de 2005 (dólar NT)
- **PAY_AMT6:** cantidad del pago anterior en abril de 2005 (dólar NT)



Figure 2: Descripción de los datos

2. Arquitectura de la Solución (Diagrama final)

Es importante destacar que la arquitectura del proyecto se mantuvo sin cambios durante la fase de desarrollo y despliegue. Esta consistencia garantizó que la ejecución del proyecto estuviera alineada con los objetivos establecidos.

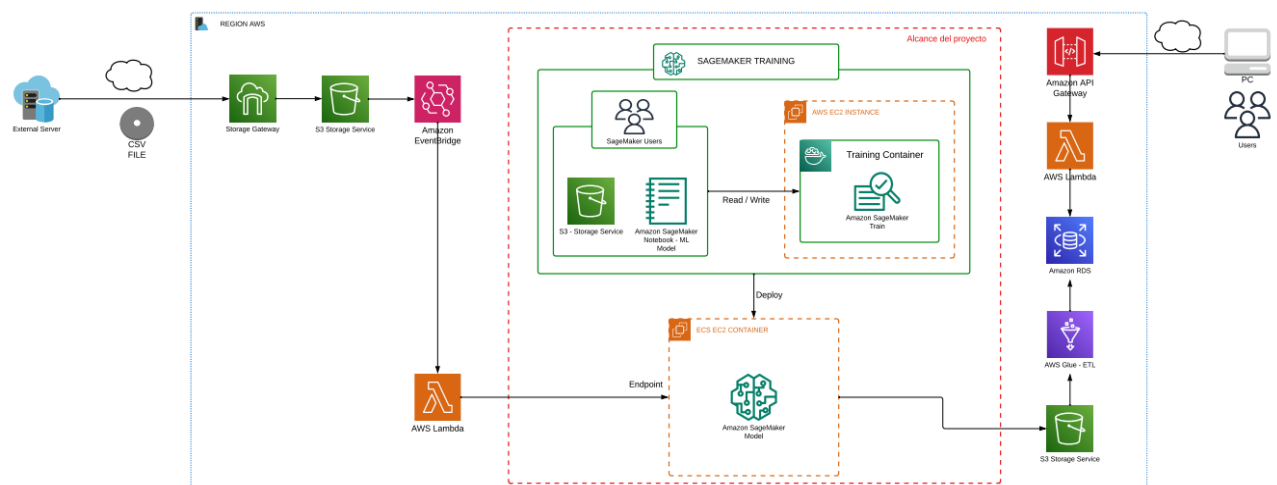


Figure 3: Arquitectura del proyecto SMI



3. Código Fuente


En esta sección, se presentará el código fuente del proyecto el cual estará anexo a las respectivas entregas, el cual está dividido en dos cuadernos para una mejor organización y comprensión.

El primer cuaderno abarca las tres partes iniciales del proyecto: la **Parte 1** se centra en el entendimiento de los datos y su limpieza, la **Parte 2** se enfoca en el modelado del modelo de Machine Learning, y la **Parte 3** corresponde a la **fase A** del despliegue del modelo.

Por otro lado, el segundo cuaderno está dedicado a la **Fase 2** del despliegue del modelo en AWS SageMaker, donde se implementan las configuraciones necesarias para asegurar un funcionamiento del modelo. Ambos cuadernos son recursos esenciales para comprender el proceso de desarrollo y las decisiones técnicas tomadas a lo largo del proyecto.

4. Interfaz Gráfica de Usuario

En esta sección, se presentará la interfaz gráfica del usuario (GUI) desarrollada hasta el momento en el proyecto, la cual permite interactuar con el modelo de Machine Learning implementado. Hasta este punto, hemos logrado mostrar el endpoint en AWS, brindando a los usuarios la capacidad de acceder al modelo de manera sencilla y efectiva.





Amazon SageMaker > Puntos de conexión

Puntos de conexión Actualizar punto de enlace Acciones Crear punto de enlace

Buscar puntos de conexión < 1 >

| | Nombre | ARN | Hora de creación | Estado | Actualización más reciente |
|-----------------------|--|---|---------------------|------------------------|----------------------------|
| <input type="radio"/> | sagemaker-xgboost-2024-09-27-23-50-856 | arn:aws:sagemaker:us-east-2:535002877348:endpoint/sagemaker-xgboost-2024-09-27-23-26-50-856 | 27/9/2024, 18:26:52 | InService | 27/9/2024, 18:29:42 |

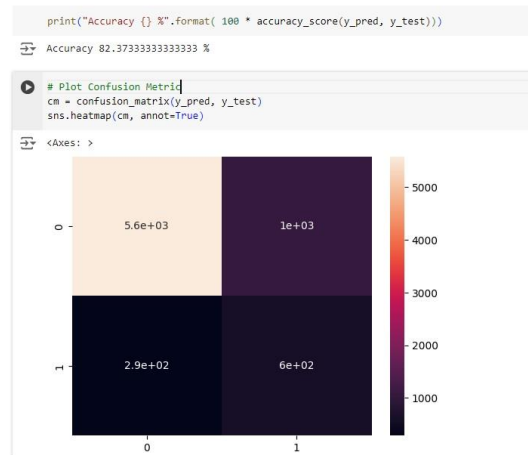


Figure 4: Interfaz gráfica de usuario - SMI