

UNIVERSIDAD AUTÓNOMA DE OCCIDENTE





# SOFTWARE MODELADOR DE IMPAGOS SMI

Angie Tatiana Aparicio Jhonn Sebastian Robles Jhon Brandon Idarraga Harold Muñoz

Esp. Analitica en Big Data **UAO** 

Septiembre, 2024



### DOCUMENTO DEFINICIÓN DEL ALCANCE

- o EDT (Estructura Desglosada de Trabajo WBS)
- o Descripción del alcance.
- Descripción del Trabajo
- Descripción de los Entregables
- Descripción de las Exclusiones

### **DOCUMENTO DE REQUISITOS**

- Backlog del Producto
- o Hoja de Recursos
- o Documentación de las Estimaciones de tiempo y costo.

### **PROTOTIPO**

- o Diagrama de Arquitectura
- o Diagrama de Flujo del proceso
- Visualización de Resultados

### PARTE 2: PLANIFICACIÓN DEL TIEMPO

- Hoja de Ruta del Proyecto.
- o Implementación en Herramienta de Software

### PARTE 3: DETERMINACIÓN DEL PRESUPUESTO

- Cuadro de Presupuesto
  - Costos por categoría y Periodo.
  - Reservas de contingencia y de gestión.
- Curva S del Costo.



PARTE 1: DEFINICIÓN DEL ALCANCE

# EDT (Estructura Desglosada de Trabajo - WBS):

Nuestra Estructura Desglosada del Trabajo (EDT) es la representación jerárquica del trabajo total que se debe realizar en el proyecto Software Modelador de Impagos - SMI. Se descompone en fases más pequeñas, que son más manejables y fáciles de controlar.

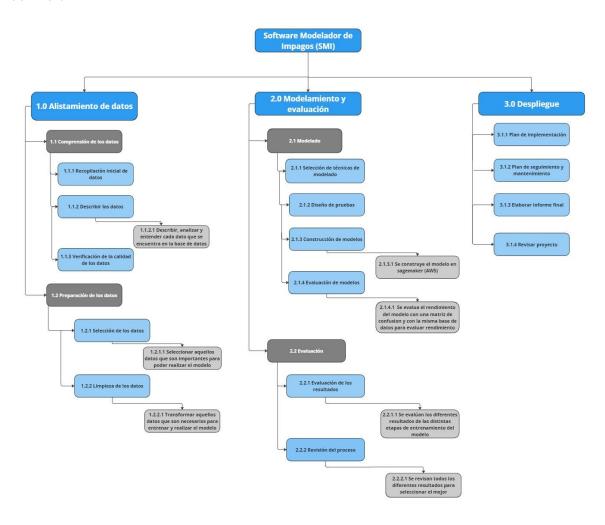


Ilustración 1. EDT - Proyecto SMI



- Descripción del alcance:
- Descripción del Trabajo:

Este proyecto se enfoca en desarrollar un modelo de Machine Learning usando Amazon SageMaker que permita predecir cuáles de los clientes del banco en Taiwán no podrán pagar sus cuotas al final del mes. El trabajo incluye las siguientes actividades clave:

- 1. **Comprensión y Preparación de Datos:** Se integrarán los datos existentes del banco, almacenados en Amazon S3, y se realizará un proceso de limpieza y preprocesamiento de datos, garantizando que la información esté lista para el entrenamiento del modelo.
- 2. **Desarrollo del Modelo Predictivo:** Se utilizará Amazon SageMaker para crear y entrenar un modelo de Machine Learning. El equipo trabajará en seleccionar el algoritmo adecuado, entrenarlo con los datos proporcionados y ajustar sus hiperparámetros para optimizar su rendimiento.
- 3. **Validación y Evaluación del Modelo:** Una vez entrenado, el modelo será validado usando un conjunto de datos de prueba.
- 4. **Despliegue del Modelo:** El modelo será desplegado en un entorno de producción utilizando Amazon SageMaker. Este entorno permitirá que el modelo sea accesible mediante un endpoint, donde podrá ser consultado en tiempo real para realizar predicciones.

### Descripción de los Entregables:

Los entregables del proyecto incluyen:

**Modelo Predictivo:** Un modelo de Machine Learning usando Amazon SageMaker entrenado y validado que predice la probabilidad de morosidad de los clientes.

**Informe de Análisis de Datos:** Documentación que detalla el proceso de recolección, limpieza y análisis de datos.

Informe de Validación del Modelo: Resultados de las pruebas y validaciones realizadas al modelo, incluyendo métricas de rendimiento con una matriz de confusión.

### Descripción de las Exclusiones:

**Integración con Sistemas Externos:** La integración con el servidor externo para la transferencia de archivos CSV a través de Storage Gateway y S3 Storage Service está fuera del alcance del proyecto.

**Automatización con EventBridge y AWS Lambda:** La configuración y manejo de flujos de trabajo automáticos que utilizan Amazon EventBridge y AWS Lambda para la activación de procesos están excluidos.

**Gestión de Almacenamiento y ETL:** Procesos de extracción, transformación y carga (ETL) mediante AWS Glue y la gestión de bases de datos relacionales con Amazon RDS no están incluidos.

**Configuración de API y Gateway:** La creación y configuración de la API para acceder al modelo a través de Amazon API Gateway está fuera del alcance.

**Postprocesamiento y Servicios Adicionales:** Servicios adicionales después de la implementación del modelo, como el almacenamiento de resultados en S3 o su uso en otros sistemas, no forman parte del proyecto.

**Análisis Financiero Detallado:** No se realizará un análisis financiero detallado del impacto del modelo en la rentabilidad del banco, más allá de las pruebas de rendimiento estándar del modelo (Matriz de confusión).

**Desarrollo de Software Personalizado:** No se contempla el desarrollo de aplicaciones o interfaces personalizadas.



### **DOCUMENTO DE REQUISITOS**

### Backlog del Producto:

El Backlog del Producto se encuentra documentado en el archivo de **Excel Plantillas Planificación\_SMI** en la hoja titulada **PBacklog**. Aquí se enumeran y priorizan todos los entregables, requisitos y los criterios de aceptación que deben ser implementados durante el proyecto, con su respectiva descripción y prioridad de desarrollo.

### Hoja de Recursos:

La Hoja de Recursos se encuentra en la hoja titulada <u>HRecursos</u> del archivo de **Excel Plantillas Planificación\_SMI**. Esta hoja incluye una lista detallada de los recursos humanos, tecnológicos y materiales necesarios para la ejecución del proyecto.

## Documentación de las Estimaciones de tiempo y costo:

Para el desarrollo del proyecto, estimamos costos tanto de tipo trabajo, como de tipo material, en cuanto al tipo trabajo, tenemos los cuatro miembros del equipo con un rol especifico y horas de dedicación asignadas, como lo mostramos en la siguiente tabla:

Rol	Horas	Salario mes (promedio en Colombia)	Calculo costo
Project Manager	62	\$ 4.500.000	(4.500.000/30)/8*62 = <b>1.162.500</b>
Machine Learning Engineer	155	\$ 8.400.000	(8.400.000/30)/8*155 = <b>5.425.000</b>
Data Engineer	124	\$ 6.000.000	(6.000.000/30)/8*124 = <b>3.100.000</b>
Data Analyst	93	\$ 4.500.000	(4.500.000/30)/8*93 = <b>1.743.750</b>

El cálculo del costo por cada rol en el proyecto tenemos el salario mes, sobre los días del mes para sacar el valor por día, luego se divide por 8 que son las horas laborales que tiene un día y se multiplica por el número de horas que dedicará cada persona.

Para estimar la contingencia, trabajamos con los mismos valores de salario y la fórmula para el cálculo del costo, pero apreciando horas adicionales por cada rol, como si el proyecto se extendiera una semana más (7 días):

Rol	Horas (contingencia)	Salario mes (promedio en Colombia)	Calculo costo
Project Manager	14	\$ 4.500.000	(4.500.000/30)/8*14 = <b>262.500</b>
Machine Learning Engineer	56	\$ 8.400.000	(8.400.000/30)/8*56 = <b>1.960.000</b>
Data Engineer	28	\$ 6.000.000	(6.000.000/30)/8*28 = <b>700.000</b>
Data Analyst	21	\$ 4.500.000	(4.500.000/30)/8*21 = <b>393.750</b>

En cuanto a los materiales que serán usados tenemos lo siguiente:

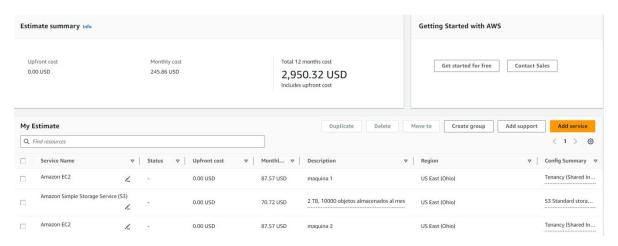


Ilustración 2. Calculadora de AWS con simulación del costo de las maquinas a utilizar

Dos máquinas virtuales (Amazon EC2) de 16 Gb de RAM, cada una por un valor de 87,57 USD que multiplicado por la TRM \$ 4.185,82 serían \$ 366.552 COP, para un total de \$733.105 COP y un Amazon Simple Storage Service (S3) de 2 TB de almacenamiento a un valor de 70,72 USD, que al pasarlo a pesos colombianos daría como resultado \$296.021, para ser usado por el tiempo de un mes cada servicio.

### **PROTOTIPO**

Este tablero es un **prototipo de entrega** que refleja el estado final de un proyecto de ML enfocado en predecir el riesgo de morosidad o impago de clientes de un banco en Taiwán en relación con sus tarjetas de crédito.

Aunque este tablero muestra un resumen gráfico de métricas clave como la cantidad de clientes, usuarios activos, transacciones y tasas de morosidad, es importante señalar que la creación de este tablero **no forma parte del alcance del proyecto actual**.

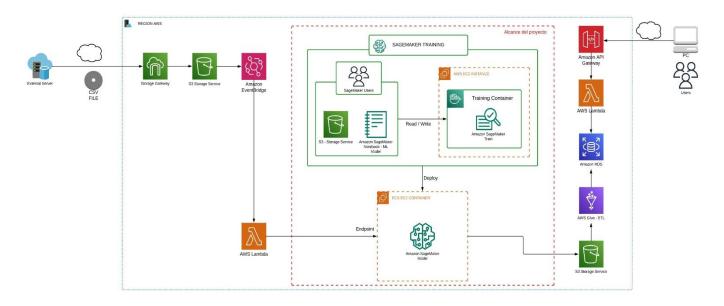
Su propósito es ilustrar cómo se podría visualizar el resultado final del proyecto completamente implementado, este tablero sirve como una propuesta de cómo los resultados podrían ser monitoreados en tiempo real mediante un informe visual completo y detallado.

Al destacar la importancia del modelo predictivo, este tablero facilita la toma de decisiones en cuanto a la recuperación de deudas y la gestión de los clientes con mayor probabilidad de caer en morosidad.



Ilustración 3. Prototipo de Dashboard del proyecto finalizado SMI





ARQUITECTURA DE LA SOLUCION

Ilustración 4. Diagrama de arquitectura SMI

Nuestra arquitectura presenta el desarrollo y despliegue de un modelo de Machine Learning (ML) para la predicción de morosidad en clientes de un banco de Taiwán. El proyecto se llevó a cabo utilizando tecnologías de la nube como **Amazon SageMaker** para el entrenamiento y despliegue del modelo.

La **arquitectura global** de la solución está diseñada para garantizar escalabilidad, seguridad y eficiencia, incluye múltiples servicios de **AWS** como **S3 Storage Service**, **AWS Lambda** y **Amazon API Gateway**. Estos componentes están interconectados para automatizar el flujo de datos desde su almacenamiento hasta el uso del modelo en tiempo real.

Sin embargo, es importante señalar que **el alcance de este proyecto está limitado exclusivamente a los componentes punteados en rojo en el esquema** de la arquitectura. Esto incluye las actividades relacionadas con el desarrollo, entrenamiento y despliegue del modelo utilizando Amazon SageMaker, excluyendo otros aspectos como la integración con sistemas externos o la gestión de APIs.

Para una visualización más detallada de la arquitectura, se puede consultar el documento titulado **Arquitectura de la Solución** que se encuentra en el PDF adjunto a este informe. Este archivo incluye una representación gráfica de todos los componentes y servicios involucrados.

### Diagrama de Flujo del proceso:

A continuación, se presenta el Diagrama de Flujo del Proceso, que ilustra de manera clara y concisa cada una de las etapas clave involucradas en el desarrollo y despliegue del modelo de Machine Learning. Este diagrama proporciona una visión global del flujo de trabajo, desde la recolección de datos hasta el uso final del modelo.

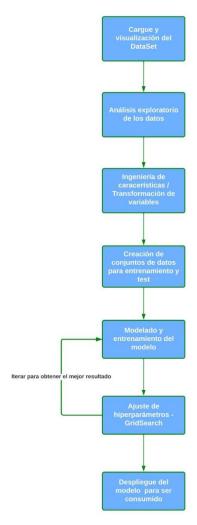


Ilustración 5. diagrama de flujo proyecto SMI



### Visualización de Resultados:

El principal entregable del proyecto es el **Modelo Predictivo** desarrollado con Amazon SageMaker, que permitirá predecir la probabilidad de que los clientes incurran en morosidad. A través de múltiples iteraciones de entrenamiento y ajuste, el modelo será optimizado para ofrecer predicciones precisas, lo cual es fundamental para la toma de decisiones del banco.

Para ello entregaremos representaciones gráficas que muestrearán el rendimiento del modelo a lo largo del tiempo, en donde destacaremos la capacidad para predecir correctamente los casos de morosidad.

El Informe de Análisis de Datos detallará el proceso de recolección y limpieza de los datos utilizados para entrenar el modelo. La calidad de los datos es clave para la precisión del modelo, por lo que este informe explicará las transformaciones realizadas y justificará las decisiones tomadas en cuanto a la manipulación de los datos.

El Informe de Validación del Modelo deberá contener los resultados en donde se demostrará la precisión y capacidad predictiva. Se entregará una matriz de confusión que muestra el número de predicciones correctas e incorrectas (verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos), permitiendo justificar el rendimiento del modelo.

# PARTE 2: PLANIFICACIÓN DEL TIEMPO

### Hoja de Ruta del Proyecto:

La Hoja de Ruta del Proyecto se encuentra detallada en el archivo de Excel Plantillas Planificación\_SMI, específicamente en la hoja titulada HRuta. En esta hoja, se presentan todas las fases del proyecto, incluyendo los objetivos de cada fase o iteración, así como también los ítems correspondientes al producto backlog.



Para la implementación y seguimiento del proyecto, se utilizó la herramienta **JIRA**, en donde se registraron las tareas, los sprints, y se dará el respectivo seguimiento al avance del proyecto. La integración de JIRA nos permitirá asegurar una organización ágil y efectiva, facilitando la colaboración entre los miembros del equipo y el control de cada etapa clave.

# PARTE 3: DETERMINACIÓN DEL PRESUPUESTO

### Cuadro de Presupuesto

- Costos por categoría y Periodo.
- Reservas de contingencia y de gestión.

El Cuadro de Presupuesto se encuentra detallada en el archivo de **Excel Plantillas Planificación\_SMI**, específicamente en la hoja titulada **Presupuesto**, el cual proporciona un desglose detallado de los costos por categoría y período, así como las reservas de contingencia y de gestión.

### Curva S del Costo.

La curva S se encuentra detallada en el archivo de **Excel Plantillas Planificación\_SMI**, específicamente en la hoja titulada **Curva S**, la cual proporciona una representación gráfica de los costos acumulados a lo largo del tiempo en función de la curva S, que ayuda a visualizar el progreso del gasto en relación con el cronograma del proyecto.