

# XGScore - Predicción unificada de riesgo crediticio

Omar García, Tabaré Merino, Fernando Lango

Ciudad de México, mayo de 2025

## 1 Press Release

Un equipo de analistas de datos ha desarrollado una solución avanzada para la gestión del riesgo crediticio, combinando algoritmos de machine learning con una arquitectura moderna en la nube utilizando servicios de **Amazon Web Services (AWS)**. Esta herramienta permite monitorear y predecir el riesgo de incumplimiento en portafolios de crédito de forma mensual, eficiente y escalable.

En un entorno financiero cada vez más competitivo, las instituciones enfrentan el reto de mantener la calidad de sus portafolios sin sacrificar su capacidad de crecimiento. La clave para lograrlo está en anticipar el riesgo. Con esa premisa, esta solución automatiza el análisis de más de 1.2 millones de registros provenientes del buró de crédito y más de 2.4 millones de registros de comportamiento de pago históricos, utilizando modelos entrenados sobre cinco dimensiones fundamentales del comportamiento crediticio.

La infraestructura de la solución está diseñada sobre la nube de AWS. Los datos son gestionados mediante **Amazon S3**, **AWS Glue** y **AWS Athena**, lo que permite automatizar su ingesta y transformación de manera robusta y segura. Para, el entrenamiento y la predicción de modelos se realizan cinco modelos independientes de **XGBoost** especializados en distintas categorías: crédito revolving, crédito a plazos, morosidad, uso de crédito y actividad reciente. Un modelo de **ensamble** combina estas salidas para generar un score integrado por cliente.

La visualización del riesgo se realiza a través de **Looker Studio**, lo que permite a los tomadores de decisiones acceder a dashboards interactivos para monitorear la evolución del riesgo agregado, identificar segmentos críticos y evaluar tendencias históricas por cohorte.

La métrica clave utilizada es **Ever30@12M**, que identifica si un cliente incurre en más de 30 días de atraso en los 12 meses siguientes. Esta definición permite focalizar la evaluación en clientes aún activos, excluyendo casos con morosidad evidente y maximizando el valor predictivo de los modelos.

## 2 Cinco preguntas alrededor del cliente

### 1. ¿Quién es el cliente?

- La institución financiera que desea monitorear y mejorar el riesgo de su portafolio de crédito.

### 2. ¿Cuál es su problema?

- No cuenta con una herramienta para visualizar la posición de riesgo del portafolio ni su tendencia histórica.

### 3. ¿Qué desea lograr el cliente?

- Responder mediante el uso de variables de buró de crédito si su perfil de clientes ha mejorado históricamente y tomar decisiones estratégicas con base en dicha información.

#### 4. ¿Qué insumos tiene el cliente?

- Base con 180 variables del buró de crédito, de enero 2022 a abril 2025 (1,289,324 registros).
- Base de comportamiento con 5 variables (2,408,872 registros).

#### 5. ¿Qué obtiene el cliente?

- Un score de riesgo mensual basado en cinco índices independientes que permiten evaluar la salud financiera del portafolio.

## 3 Boceto de la solución final

La solución propuesta para el análisis de riesgo de un portafolio de crédito mediante variables de buró está diseñada para ejecutarse de forma mensual y automatizada sobre una arquitectura escalable basada en servicios de AWS (Amazon Web Services). Su objetivo es calcular un score de riesgo que refleje la probabilidad de que un cliente incurra en más de 30 días de atraso en los próximos 12 meses (Ever30@12M). A continuación, se describe el flujo detallado de la solución y su soporte técnico.

### 3.1 Arquitectura General

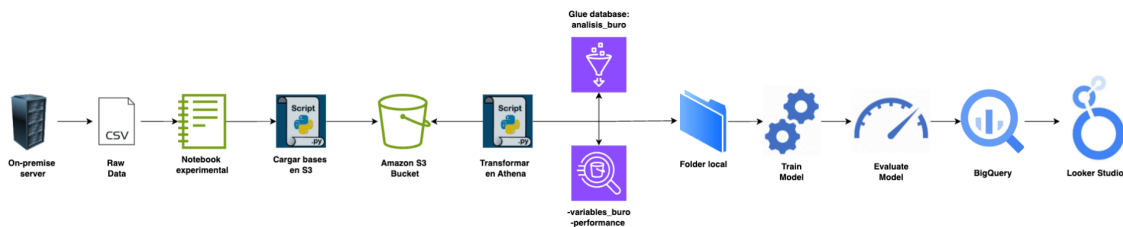


Figure 1: Arquitectura de la solución

### 3.2 Ingesta de datos

Los datos provienen de dos fuentes principales: - Base de buró de crédito con 180 variables por cliente, con frecuencia mensual (enero 2022 - abril 2025), con más de 1.28 millones de registros. - Base de comportamiento de pago con 5 variables clave, misma frecuencia, con más de 2.4 millones de registros.

Estos datos son cargados en formato CSV en un bucket de Amazon S3. Se automatiza esta carga mediante tareas periódicas con AWS Glue.

### 3.3 Preprocesamiento y feature engineering

Mediante un job de AWS Glue, se realiza la transformación mensual de los datos: - Unión de bases de buró y comportamiento por ID y mes de información - Generación de etiquetas de riesgo Ever30@12M (variable binaria)

El resultado es almacenado como tabla analítica en Amazon Athena.

### 3.4 Modelado por categoría de riesgo

Se construyen cinco modelos independientes utilizando XGBoost , cada uno enfocado en una dimensión del riesgo crediticio: - Revolving: comportamiento en líneas de crédito rotativas - Installment: comportamiento en productos de pagos fijos - Credit Activity: número de aperturas, cierres y consultas recientes - Delinquency: historial de morosidad - Credit Utilization: uso relativo del crédito disponible

Cada modelo se entrena usando los datos de la ventana enero 2022 - abril 2024.

### 3.5 Ensamble y generación de score global

Los scores de los cinco modelos se combinan mediante un modelo meta de ensamble. Esto genera un score único y mensual para cada cliente.

### 3.6 Visualización y análisis de resultados

Los resultados se exportan en formato tabular, se llevan a bigquery y se visualizan en looker studio. Se pueden construir dashboards con: - Evolución histórica del score por cliente - Score promedio por cohorte temporal o segmento - Porcentaje del portafolio con Ever30@12M estimado alto

### 3.7 Automatización y mantenimiento

Toda la solución puede ser orquestada mensualmente mediante scripts. Las métricas de rendimiento del modelo (precisión, F1, AUC) se recalculan periódicamente para validar su estabilidad.

## 4 FAQ: Preguntas Frecuentes

#### 1. ¿Qué tecnologías y algoritmos se utilizaron?

Se utilizaron modelos supervisados basados en **XGBoost**, uno para cada una de las cinco dimensiones del comportamiento crediticio. Posteriormente, estos índices se integran en un modelo de **ensamble**, generando un score de riesgo mensual por cliente.

#### 2. ¿Cómo se aplica la solución en la operación mensual de la empresa?

La solución se ejecuta mensualmente. Utiliza los datos más recientes del buró de crédito y del comportamiento de pago para calcular nuevos índices y actualizar el score global del cliente. Esto permite monitorear la evolución del portafolio de forma continua.

#### 3. ¿Qué tipo de datos se requieren como insumo?

- Registros mensuales del buró de crédito con 180 variables por cliente.
- Registros mensuales del comportamiento de pago con 5 variables clave.

Ambos conjuntos cubren el periodo de enero 2022 a abril 2025.

#### 4. ¿Qué métrica se usa para medir el riesgo del portafolio?

Se define como la probabilidad de que un cliente incurra en más de 30 días de atraso en los siguientes 12 meses. Esta condición se expresa mediante la variable binaria **Ever30@12M**, usada como etiqueta para entrenar los modelos.

#### 5. ¿Cuáles son los principales inputs y outputs del sistema?

##### Inputs:

- Datos mensuales del buró de crédito (180 variables)
- Datos mensuales del comportamiento de pago (5 variables)

##### Outputs:

- Cinco índices de riesgo por cliente

- Score mensual agregado por cliente

6. **¿Cómo se interpreta y utiliza el score generado?**

El score mensual refleja el riesgo de incumplimiento de un cliente. Puede ser usado para:

- Aprobar o rechazar solicitudes de crédito con mayor precisión.
- Priorizar acciones de cobranza preventiva.
- Analizar la evolución histórica del portafolio.

7. **¿Qué tipo de analítica se utilizó?**

- Modelos de aprendizaje automático supervisado (XGBoost).
- Ensamble de modelos.
- Análisis de evolución temporal.

8. **¿La solución contempla visualización de resultados?**

Sí. Se puede dar seguimiento al score mensual del portafolio y de cada cliente, así como analizar:

- El score promedio por cohorte o segmento.
- Tendencias temporales del riesgo agregado.
- Distribución del riesgo por categoría.

9. **¿Qué tan generalizable es esta solución?**

El modelo fue entrenado con datos entre enero 2022 y abril 2024. Puede aplicarse a clientes con historial suficiente para cumplir la ventana de 12 meses usada para calcular Ever30@12M.

10. **¿Se incluyó una estimación de costos?**

Sí, el costo anual de evaluar al portafolio de manera mensual con los servicios antes nombrados es de menos de 25 dólares.