**CIENTIFICO DE DATOS**

**SISTECREDITO SAS**

**Proceso de Selección**

**Nombre Completo:**

**Numero Identificación:**

**Observaciones:**

* Use este documento para incluir la mayor cantidad de la información textual y gráfica que responda a las preguntas.
* Es posible que requiera utilizar herramientas externas. Incluya los enlaces de los artefactos externos, o los archivos adjuntos correspondientes dentro del correo. Se recomienda que pueda respaldar los distintos hallazgos con su debido soporte.
* Los siguientes casos, corresponden a la mezcla de información cierta y ficticia, construidos para fines de medir ciertos conocimientos y habilidades requeridos para el cargo en mención. Los datos que se proveen para el ejercicio han sido anonimizados y no representan de ninguna forma a los clientes de Sistecredito SAS.
* No hay una extensión mínima o máxima.
* Guarde este documento en formato PDF, cuyo nombre debe tener la siguiente estructura: “<primer\_nombre>\_<primer\_apellido>\_casos”. Por ejemplo: ricardo\_sanchez\_casos.
* Para cualquier otro archivo, utilice la siguiente estructura: “<primer\_ nombre>\_<primer\_ apellido>\_<alias\_archivo>”. Por ejemplo: ricardo\_sanchez\_depuracion\_datos.
* Envíe todos los documentos y soportes en un único correo, con el asunto: “Nombre del aspirante\_CientificoDatos”, a los siguientes correos: [lmendoza@sistecredito.com](mailto:lmendoza@sistecredito.com), [jomunoz@sistecredito.com](mailto:jomunoz@sistecredito.com), [yguzman@sistecredito.com](mailto:yguzman@sistecredito.com) . En el correo, explique brevemente el contenido de los archivos adicionales, si llega a tenerlos.
* Si tiene alguna duda o solicitud al respecto, por favor comunicarse a los correos: [lmendoza@sistecredito.com](mailto:lmendoza@sistecredito.com); [jomunoz@sistecredito.com](mailto:jomunoz@sistecredito.com), yguzman@sistecredito.com.

**Prueba científico de Datos.**

1. ¿Usted ha tenido experiencia en Ciencia de datos? Si/No. En caso de ser afirmativo, explique cuál ha sido, en qué proyectos ha trabajado y que logros ha obtenido.
2. De todo el scope de ciencia de datos, cual considera que es su fuerte? En que temas considera que tiene debilidades técnicas y/o de negocio?

# **Prueba técnica para ciencia de datos: Riesgo crediticio**

## Contexto:

El banco ha identificado una problemática relacionada con el deterioro de la cartera en compras realizadas en el comercio electrónico internacional 'Temu'. Si bien los clientes utilizan sus tarjetas de crédito en diferentes establecimientos, las compras en 'Temu' han mostrado un mayor nivel de incumplimiento frente al promedio. Por este motivo, surge la necesidad de desarrollar un modelo que permita predecir si un cliente incurrirá o no en mora al usar su tarjeta en dicho comercio. Esta iniciativa busca minimizar riesgos financieros y fortalecer las políticas de otorgamiento de crédito.

## Datos del reto

Se suministra un archivo Excel compuesto por dos hojas:

- Hoja 'DataFramePrueba': contiene el set de datos con información de clientes que han usado su tarjeta de crédito en 'Temu'.

- Hoja 'Diccionario': incluye la descripción de cada variable.

Cada registro corresponde a un cliente y su historial de comportamiento de crédito. Dentro del dataset se encuentran clientes que ya tenían historial previo con la tarjeta, y clientes cuyo primer uso fue directamente en 'Temu'. Esto se identifica mediante la variable 'NumeroCreditosGPrevius', donde un valor de cero indica que el primer uso de la tarjeta fue en 'Temu'. La variable DiasMora representa los días de mora que tiene el cliente actuales, no cuando hizo uso de la tarjeta por lo tanto es un error tomarla como variable predictora.

# **Objetivos y Criterios de Evaluación**

## Análisis Exploratorio de Datos (EDA)

Objetivo: Evaluar la capacidad para explorar y comprender el dataset.  
  
Qué buscamos:  
- Identificación de patrones, valores faltantes, outliers y correlaciones.  
- Análisis descriptivo de las variables más relevantes.  
- Uso de visualizaciones para soportar los hallazgos.  
  
Criterios de Evaluación: Claridad del análisis, profundidad en las conclusiones y justificación lógica.

## Limpieza y Preprocesamiento de Datos

Objetivo: Evaluar la habilidad para preparar y estructurar los datos.  
  
Qué buscamos:  
- Manejo adecuado de valores faltantes, duplicados y errores.  
- Documentación clara de las decisiones de limpieza.  
  
Criterios de Evaluación: Rigurosidad, consistencia y justificación de cada paso.

## Feature Engineering

Objetivo: Medir la capacidad para generar nuevas variables que fortalezcan el modelo.  
  
Qué buscamos:  
- Creación de variables relevantes a partir de la información disponible.  
- Explicación del impacto esperado de estas variables.  
  
Criterios de Evaluación: Creatividad, pertinencia y claridad en la justificación.

## Modelado Predictivo

Objetivo: Evaluar la capacidad para seleccionar, construir y evaluar modelos predictivos.  
  
Qué buscamos:  
- Selección adecuada de algoritmos de clasificación.  
- Implementación de técnicas de validación y ajuste de hiperparámetros.  
- Interpretación de resultados y relevancia de variables.  
  
Criterios de Evaluación: Adecuación de los modelos, proceso de validación y claridad en la interpretación.

## Conocimientos Técnicos en Machine Learning/Deep Learning

Objetivo: Medir el nivel de dominio en algoritmos de aprendizaje automático.  
  
Qué buscamos:  
- Implementación de algoritmos adecuados al problema.  
- Capacidad de optimización y ajuste fino de modelos.  
  
Criterios de Evaluación: Razonamiento detrás de la selección de algoritmos y precisión en los resultados.

## Evaluación de Métricas y Resultados

Objetivo: Evaluar la correcta elección e interpretación de métricas.  
  
Qué buscamos:  
- Selección de métricas relevantes como precisión, recall, F1-score, ROC-AUC.  
- Explicación clara de los resultados en términos de negocio.  
  
Criterios de Evaluación: Pertinencia de métricas y claridad en la interpretación.

## Documentación y Comunicación

Objetivo: Evaluar la capacidad de documentar y comunicar los hallazgos.  
  
Qué buscamos:  
- Código claro y comentado.  
- Presentación ejecutiva que resuma el análisis, modelos y recomendaciones.  
  
Criterios de Evaluación: Organización, claridad y relevancia en la comunicación.

Objetivo: Evaluar la capacidad del candidato para empaquetar y exponer el modelo en un entorno reproducible y utilizable por otros equipos.

## Construcción de una API Funcional y Reproducible

Qué buscamos:  
- Desarrollo de una API REST (ej. con FastAPI o Flask) que permita consultar el modelo predictivo.  
- Entrega de instrucciones claras para la ejecución de la API (requirements.txt, Dockerfile o notebooks ejecutables).  
- Capacidad para manejar inputs reales y devolver predicciones en un formato estandarizado (ejemplo: JSON).  
  
Criterios de Evaluación: Funcionamiento correcto, claridad en la documentación, reproducibilidad y buenas prácticas de despliegue.

## Habilidades en IA Generativa

Objetivo: Medir la creatividad y capacidad del candidato para aplicar técnicas de IA Generativa en un contexto de negocio.  
  
Ejemplos de formas en las que puede demostrarlo:  
1. Explicabilidad con LLMs: Usar un modelo generativo (ej. GPT) para traducir la salida del modelo predictivo a un lenguaje entendible por negocio.  
2. Generación de reportes automáticos: Construir prompts o plantillas que automaticen la creación de informes ejecutivos a partir de los resultados.  
3. RAG (Retrieval Augmented Generation): Implementar un sistema de consulta en lenguaje natural sobre el dataset o el diccionario de variables.  
4. Generación de hipótesis o insights: Usar un LLM para sugerir nuevas variables, hipótesis de negocio o posibles mejoras al modelo.  
5. Chatbot demostrativo: Integrar el modelo predictivo con un asistente conversacional que permita simular interacciones de negocio.  
6. Explicación de sesgos o riesgos: Utilizar un modelo generativo para analizar y redactar posibles sesgos en los datos y proponer estrategias de mitigación.  
  
Criterios de Evaluación: Creatividad en la aplicación, pertinencia al caso de negocio, claridad en la integración y capacidad para traducir conceptos técnicos a valor para la organización.

Siéntete libre de sorprendernos 😊

## Condiciones de Entrega

El candidato deberá entregar:  
- Un notebook con el análisis, exploración, modelado y conclusiones.  
- Una presentación ejecutiva clara y concisa dirigida a un público de negocio.  
  
La entrega debe incluir justificación de las decisiones, limitaciones identificadas y propuestas de mejora. El archivo final debe ser guardado en formato PDF con la nomenclatura '<nombre\_apellido\_casos>'. Los archivos adicionales deben seguir el mismo formato de nomenclatura.   
  
Los entregables deberán enviarse en un único correo con el asunto: 'Nombre del aspirante\_CientificoDatos'. El correo debe incluir una breve explicación de los archivos enviados.  
  
Una vez concluida la prueba, el candidato se compromete a eliminar los datos suministrados.