

# Sprint 3: Proyecto de ML para incrementar el gasto anual de clientes en la tienda



## Introducción

Se trata de una tienda exclusiva especializada en la confección y venta de ropa a medida. Se destaca por ofrecer consultorías altamente personalizadas. Los clientes visitan la tienda para recibir asesoramiento directo de estilistas expertos que ayudan a crear prendas únicas que se ajustan perfectamente a sus preferencias y medidas. Después de la sesión de consultoría, los clientes pueden hacer pedidos de ropa a través de una aplicación móvil o del sitio web de la empresa.

## Objetivos del Proyecto

1. ¿Cuáles son los objetivos del negocio?  
Aumentar las ventas evaluando dónde concentrar los esfuerzos: mejorando la experiencia de los clientes en el sitio web o en la aplicación móvil.
2. ¿Qué decisiones o procesos específicos desea mejorar o automatizar con ML?  
Se busca optimizar las decisiones empresariales relacionadas con la experiencia del cliente y automatizar la predicción del gasto anual de cada cliente.
3. ¿Se podría resolver el problema de manera no automatizada?  
Aunque la predicción del gasto anual de cada cliente se puede realizar utilizando hojas de cálculo o software estadístico, el uso de machine learning permitirá automatizar este proceso mediante entrenamientos programados que incorporen rápidamente las transacciones diarias sin necesidad de intervención manual.

## Metodología Propuesta

4. ¿Cuál es el algoritmo de Machine Learning más adecuado para resolver este problema? ¿Cómo justifica la elección de este algoritmo? ¿Qué métricas de evaluación se utilizarán para medir el rendimiento del modelo?

**XGBoost** es un algoritmo de boosting que ha demostrado ser altamente eficaz en problemas de predicción, especialmente en escenarios donde se busca un equilibrio entre precisión y eficiencia. Este algoritmo tiene la capacidad de modelar relaciones complejas entre las variables, lo que lo hace ideal para datos que incluyen interacciones no lineales, como es probable que ocurra en un entorno de comportamiento de clientes (historial de compras, interacción en la web o la aplicación, estado de membresía, etc.).

## Justificación de la Elección:

1. **Manejo de Datos Complejos:** XGBoost es capaz de capturar patrones complejos y no lineales que pueden no ser fácilmente identificables por algoritmos más simples, como la regresión lineal.
2. **Eficiencia:** Aunque es un algoritmo avanzado, XGBoost es conocido por su alta eficiencia y velocidad de entrenamiento, lo que es ideal para proyectos en los que se deben procesar grandes volúmenes de datos de manera ágil.
3. **Robustez ante el Ruido:** XGBoost es menos propenso al sobreajuste (overfitting) gracias a sus mecanismos de regularización y control del sobreajuste, lo cual es crítico cuando se trabaja con datos de clientes que pueden contener ruido o información irrelevante.
4. **Escalabilidad:** Si en el futuro los datos crecen o se integran más fuentes de información, XGBoost es altamente escalable, por lo que podrá seguir funcionando bien a medida que se añadan más datos.

## Métricas de Evaluación:

Para medir el rendimiento del modelo de XGBoost en este contexto, se utilizarían las siguientes métricas:

1. **MAE (Error Absoluto Medio):** Mide la diferencia promedio entre las predicciones y los valores reales, dándole igual peso a todos los errores. Es útil para obtener una interpretación sencilla de cuán desviado está el modelo, en promedio.
2. **R<sup>2</sup> (Coeficiente de Determinación):** Esta métrica ayuda a determinar qué tan bien el modelo explica la variabilidad del gasto anual de los clientes. Un valor cercano a 1 indica que el modelo se ajusta bien a los datos.

Es clave validar el modelo con un conjunto de datos de prueba para asegurarse de que no solo funcione bien en los datos de entrenamiento, sino que generalice correctamente a nuevos datos. Esto asegurará que el modelo predice adecuadamente el gasto futuro de los clientes y no simplemente memoriza los datos históricos.

## Datos Disponibles

5. ¿Qué datos están disponibles para abordar este problema?

La tienda cuenta con un conjunto de datos actualizado que incluye información identificativa de cada cliente, la suma anual que ha gastado en la tienda, el tiempo dedicado a interactuar tanto en el sitio web como en la aplicación móvil, y el estado de membresía.

## Métrica de Éxito

6. ¿Cuál es la métrica de éxito para este proyecto?

Aumento en el Gasto Anual Promedio por Cliente. Esta métrica reflejará directamente la efectividad del modelo en mejorar las decisiones de la empresa.

## Responsabilidades Éticas y Sociales

7. ¿Qué responsabilidades éticas y sociales es importante tener en cuenta?

Al predecir el gasto anual de los clientes mediante Machine Learning, es fundamental considerar los siguientes aspectos éticos y sociales:

1. **Privacidad y Protección de Datos:** Los datos deben cumplir con las leyes o normativas de privacidad. Es crucial anonimizar o pseudonimizar la información personal y utilizarla solo con el consentimiento explícito de los clientes.
2. **Transparencia del Modelo:** El modelo debe ser explicable y comprensible. Los clientes deben saber cómo se utiliza su información y por qué se toman decisiones específicas.
3. **No Discriminación:** Se debe evitar que el modelo introduzca sesgos que puedan perjudicar a ciertos grupos de clientes. Es vital evaluar si las variables seleccionadas generan alguna forma de discriminación y corregirlo si es necesario.
4. **Impacto a Largo Plazo:** El modelo podría afectar la experiencia en tienda física. Es clave supervisar cómo los cambios impulsados por el ML impactan a la empresa y a los clientes en el tiempo.
5. **Confianza del Cliente:** El modelo debe alinearse con el valor central de la tienda: ofrecer una experiencia personalizada. Los clientes deben sentir que sus interacciones son valiosas y respetadas.

Considerar estos principios garantizará un modelo justo, ético y alineado con los valores de la tienda y las expectativas de los clientes.