

# 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 ajusten perfectamente a sus preferencias y medidas. Después de la sesión de consultoría, los clientes pueden realizar 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 se desean 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?

Si bien la predicción del gasto anual de cada cliente puede realizarse utilizando hojas de cálculo o software estadísticos, 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 se justifica la elección de este algoritmo? ¿Qué métricas de evaluación se utilizarán para medir el rendimiento del modelo?

El algoritmo más adecuado para resolver este problema es el **Regresor de Bosques Aleatorios (Random Forest Regressor)**. Este algoritmo es eficaz para predecir valores numéricos, como el gasto anual de los clientes, porque maneja bien datos complejos y relaciones no lineales. Además, es robusto frente a valores atípicos y reduce el riesgo de sobreajuste (overfitting) al promediar los resultados de múltiples árboles de decisión y puede manejar variables categóricas como es el caso del estado de la membresía.

**Las métricas de evaluación que se utilizarán son:**

**Error Cuadrático Medio (MSE | Mean squared error):** Para medir la diferencia entre los valores predichos y los reales.

**R<sup>2</sup> (Coeficiente de Determinación):** Para evaluar qué tan bien el modelo explica la variabilidad del gasto anual.

**Teniendo en cuenta que las features disponibles incluyen:**

- Identificación del cliente (ID del cliente)
- Suma anual gastada en la tienda
- Tiempo dedicado a interactuar en el sitio web y la aplicación móvil
- Estado de membresía

**La justificación de que el Regresor de Bosques Aleatorios (Random Forest Regressor) es una excelente elección recae en varias razones:**

- **Capacidad de manejar tanto variables numéricas como categóricas:** Random Forest puede trabajar fácilmente con datos numéricos como el tiempo de interacción y la suma anual gastada, así como con variables categóricas como el estado de membresía. No requiere un extenso preprocesamiento o transformación de los datos categóricos, ya que los árboles de decisión dentro del bosque los manejan de manera natural.

- **Reducción del riesgo de sobreajuste (overfitting):** Dado que el modelo utiliza múltiples árboles de decisión (cada uno entrenado con diferentes subconjuntos de los datos), promedia los resultados, lo que reduce el riesgo de que un solo árbol aprenda demasiado los detalles específicos del conjunto de datos (sobreajuste) y pierda capacidad de generalización.
- **Tratamiento de relaciones no lineales:** Las interacciones entre el tiempo en la web/aplicación y el estado de membresía podrían tener efectos complejos en el gasto anual de los clientes. Random Forest maneja bien estas relaciones no lineales, que serían difíciles de capturar con modelos más simples como la regresión lineal.
- **Robustez frente a datos faltantes o ruidosos:** Si hay datos incompletos o valores atípicos en las interacciones de los clientes o el historial de compras, Random Forest puede ser más tolerante y aun así ofrecer buenas predicciones sin necesidad de eliminar o imputar datos de manera intensiva.
- **Importancia de las características:** Random Forest ofrece una forma de medir la importancia de las características, lo que ayuda a identificar cuáles de las variables (por ejemplo, tiempo en la aplicación, estado de membresía, etc.) tienen un mayor impacto en el gasto anual de los clientes. Esta capacidad de análisis es muy útil para la empresa, ya que permite enfocar los esfuerzos en optimizar esas áreas específicas.

## 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 reflejaría 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?

**Privacidad y protección de datos:** El modelo utilizará datos personales e identificativos de los clientes, como sus interacciones en la web y el historial de compras. Es esencial garantizar el cumplimiento de las leyes de protección de datos y asegurar que los datos sean almacenados y procesados de manera segura, evitando cualquier filtración o uso indebido de la información personal.

**Transparencia en las decisiones automatizadas:** Es importante que las predicciones sobre el gasto de los clientes no sean utilizadas de manera discriminatoria o para limitar las oportunidades de algunos clientes (por ejemplo, ofreciendo mejores ofertas solo a quienes gastan más). El modelo debe ser transparente y auditable, de forma que se pueda explicar cómo se toman las decisiones y evitar sesgos que perjudiquen a ciertos grupos de clientes.

**Evitar el sesgo algorítmico:** Asegurarse de que el modelo no favorezca a un tipo específico de cliente (por ejemplo, en función de su estado de membresía) a través de la segmentación. El sesgo en el modelo podría generar desigualdad en el trato y experiencia de compra de los clientes.

**Impacto social:** Aunque el objetivo es mejorar la automatización, debe evitarse que la implementación del ML afecte negativamente el trabajo humano, especialmente el de los estilistas que proporcionan una experiencia altamente personalizada. La tecnología debe complementar, no reemplazar, el valor humano en la experiencia de compra.

**Diego Lagos**

**Business Intelligence Analyst**

[Linkedin](#)

+34 632 67 31 73

[diegoandres0905@gmail.com](mailto:diegoandres0905@gmail.com)