

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 piezas únicas que se ajusten perfectamente a sus preferencias y medidas. Después de la sesión de consultoría, los clientes pueden hacer pedidos de ropa mediante 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 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 se puede realizar 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 justifican la elección de este algoritmo? ¿Qué métricas de evaluación se utilizarán para medir el rendimiento del modelo?*

Dado que se quiere predecir el gasto anual de cada cliente (variable continua), lo más acertado sería utilizar un algoritmo de regresión. Además, debido a la complejidad innata de los datos, sería útil probar algoritmos que consideren posibles relaciones no lineales entre variables. Utilizaría los siguientes algoritmos:

- **Regresión polinómica:** es un modelo ligeramente más complejo que la regresión lineal, y serviría para tener una primera visión del patrón de las variables; no obstante, habría que vigilar de no sobreajustar el modelo.

- **Random Forest Regressor:** este modelo se caracteriza por promediar las predicciones de múltiples árboles de decisión, ofreciendo un modelo más robusto y con mejor capacidad de generalización en nuevos datos.

Para evaluar el rendimiento del modelo, utilizaría las siguientes métricas:

- **Error absoluto medio (MAE):** para evaluar la magnitud del error, sin penalizar en exceso los valores atípicos o errores extremos.
- **Raíz del Error Cuadrático Medio (RMSE):** para comparar la magnitud del error de los dos modelos.
- **Coeficiente de Determinación (R2):** para ver si el modelo ajusta bien o no los datos en términos de la variabilidad en la variable de respuesta. Si el coeficiente es bajo, podría ser indicativo de que faltan variables en el modelo que impactan en el gasto anual.

Para poder validar correctamente estos modelos, también sería interesante realizar una validación cruzada para asegurar que el modelo tiene un buen desempeño en diferentes particiones de los datos, minimizando la probabilidad de sobreajuste.

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 Medio 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?*

Los aspectos éticos clave a tener en cuenta para este proyecto son:

- **Transparencia:** hay que explicar a los clientes de manera clara y comprensible cómo se utilizan sus datos, cómo funciona el modelo para hacer predicciones y cómo les afectan los resultados obtenidos.
- **Privacidad y seguridad de los datos:** al tratar con datos personales de los clientes (hábitos de compra, tiempo de interacción y estado membresía), hay que tener políticas bien definidas para protegerlos de un uso indebido.
- **Sesgo:** dada la exclusividad de la tienda, es posible que no se hayan utilizado los datos de ciertos grupos demográficos, por lo que hay que

asegurar que las recomendaciones y predicciones se realicen en base a factores legítimos y relevantes.

- **Equidad:** con relación al punto anterior, hay que asegurar que todos los clientes tengan acceso a los servicios de la tienda sin importar su grupo demográfico, evitando cualquier tipo de discriminación.