

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

¿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.

¿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.

¿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.

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

El algoritmo que considero utilizar es el de Regresión Lineal Múltiple (desde ahora RLM); al buscarse predecir un valor numérico y basándonos en diversos factores (tiempo de interacción de la web como de la app, estado de miembro, la suma anual gastada) un modelo de regresión basado en este algoritmo parece adecuado.

Aunque una ampliación de los datos podría hacer factible un algoritmo como la regresión de árboles de decisión, el uso de la RLM nos permite analizar como las variables independientes (tiempo en web/app, estado miembro, etc.) influyen de forma conjunta en una variable dependiente (el gasto anual).

Si se asume que existe una relación lineal entre estas variables predictoras y el gasto anual, la RLM puede modelar esta relación de manera efectiva. Los coeficientes de la regresión nos indican la importancia relativa de cada variable predictora en la predicción del gasto anual,

permitiéndonos comprender qué factores tienen mayor influencia en el comportamiento de los clientes tanto de la web como de la app.

Por ejemplo, mediante este algoritmo se podría saber si después de una mejora en la navegación de la página web y en la app se puede monitorizar la profundidad de las páginas vistas, el tiempo en el sitio, la tasa de rebote y su comportamiento de compra; esto también puede formar parte del proceso de poder predecir su gasto anual como del esfuerzo de la mejora de las herramientas web de la tienda.

Métricas para evaluar el rendimiento del modelo

En este caso tendría en consideración utilizar estas 4 métricas para poder evaluar el modelo:

Error Cuadrático Medio (MSE): Mide la media de los cuadrados de los errores. Es una métrica popular debido a su facilidad de cálculo y diferenciabilidad. Sin embargo, penaliza fuertemente los grandes errores, lo que puede ser relevante en nuestro caso si queremos evitar subestimar gastos elevados.

Raíz del Error Cuadrático Medio (RMSE): Es la raíz cuadrada del MSE y tiene la ventaja de estar en las mismas unidades que la variable objetivo (gasto). Esto facilita la interpretación de los resultados.

Error Absoluto Medio (MAE): Calcula el promedio del valor absoluto de los errores. Es menos sensible a valores atípicos que el MSE y proporciona una medida más robusta de la precisión del modelo.

Coefficiente de Determinación (R^2): Indica la proporción de la variabilidad de la variable dependiente (gasto) que es explicada por el modelo. Un valor de R^2 cercano a 1 indica un buen ajuste del modelo.

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

Al desarrollar y utilizar modelos de Machine Learning para predecir el gasto de los clientes y evaluar la eficiencia en el uso de las herramientas como la página web de la tienda o su app, es fundamental adoptar una perspectiva ética y social.

Esto implica respetar la privacidad de los datos, garantizar la transparencia, evitar la discriminación y utilizar los datos de manera responsable. Al hacerlo, se puede construir una relación de confianza con los clientes y contribuir a un desarrollo tecnológico más justo y equitativo.

Las consideraciones que pueden tenerse en cuenta para cumplir con ello pueden incluir, entre otros:

Privacidad de Datos

- **Consentimiento informado:** Se debe de obtener el consentimiento explícito de los clientes y ser informados sobre cómo se utilizarán sus datos y qué medidas se toman para protegerlos.

- **Minimización de datos:** Solo se deben recopilar y utilizar los datos estrictamente necesarios para alcanzar los objetivos del proyecto.
- **Seguridad de los datos:** Implementar medidas de seguridad robustas para proteger los datos de accesos no autorizados, pérdidas o alteraciones.
- **Anonimización y seudonimización:** En la medida de lo posible, los datos personales deben ser anonimizados o seudonimizados para proteger la identidad de los clientes.

Transparencia y Explicabilidad:

- **Transparencia en el uso de datos:** Los clientes deben ser informados de cómo se utilizan sus datos para personalizar la experiencia de compra.
- **Explicabilidad de los modelos:** Los modelos de ML pueden ser complejos, es importante poder explicar de manera sencilla cómo se han llegado a las predicciones.

No discriminación:

- **Evitar sesgos:** Los modelos de ML pueden perpetuar sesgos presentes en los datos de entrenamiento. Es fundamental analizar los datos para identificar y mitigar estos sesgos.
- **Trato equitativo:** Los modelos deben tratar a todos los clientes de manera justa y equitativa, independientemente de su género, raza, edad u otras características.

Uso responsable de los datos:

- **Evitar la manipulación:** Los datos no deben utilizarse para manipular o engañar a los clientes.
- **Cumplimiento de la normativa:** Asegurarse de cumplir con la normativa vigente en materia de protección de datos (en el caso de España el RGPD).