

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 consultorias altamente personalizadas. Los clientes visitan la tienda para recibir asesoramiento directo de estilistas expertos que ayuden 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. ¿ Cuales son los objetivos del negocio ?

Aumentar las ventas evaluando donde concentrar los esfuerzos: mejorando la experiencia de los clientes en el sitio web o la aplicación móvil.

2. ¿ Que 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 hacer 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. ¿ Cual es el algoritmo de Machine Learning más adecuado para resolver este problema ? ¿ Como justifica la elección de este algoritmo ? ¿ Que métricas de evaluación se utilizarán para medir el rendimiento del modelo ?

Tanto las características ("Avg. Session Length", "Time on App", "Time on Website" y "Length of Membership") como la etiqueta ("Yearly Amount Spent"), son valores continuos. Esto es, su tipo de medida es una razón, según la escala de Stevens. Suponiendo que exista una relación entre las variables, la variable de salida, "Yearly Amount Spent", crecerá o decrecerá con las diferentes variables de entrada, como "Avg. Session Length". Dado que son variables continuas y que suponemos una relación entre ellas, el algoritmo de elección sería la **regresión lineal**.

Por tanto, se realizará un análisis de **regresión lineal** en los datos. Si la, así llamada "**r**" de **pearson**, nos da un valor razonable para la linealidad de los datos, escogeremos este modelo. De lo contrario, deberemos analizar los datos por métodos no lineales. Si incluso éstos fallan, deberemos utilizar un método de agrupación como K-means.

Dado que se ha escogido la **regresión lineal** como modelo, las métricas más apropiadas serían el "Error Cuadrático Medio" o "MSE", su raíz cuadrada o "RMSE", o bien el "coeficiente de determinación R²".

Datos Disponibles.

5. ¿ Que 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 la duración del alta.

Métrica de Éxito.

6. ¿ Cual 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. ¿ Que responsabilidades éticas y sociales es importante tener en cuenta ?

De cara a los principios éticos aplicados a la IA, se definen cuatro roles: Gobiernos, empresas, desarrolladores y usuarios.

En calidad de **desarrolladores** tendremos en cuenta la privacidad, seguridad, equidad y transparencia durante el diseño e implementación del proyecto.

En cuanto a la **privacidad**, en los datos a considerar aparecen los emails y direcciones de los usuarios que serán tratados como privados.

En cuanto a la **seguridad**, el proyecto no trata con materias relacionadas con la seguridad de personas o equipos, a diferencia de aplicaciones como la conducción autónoma.

En cuanto a la **equidad**, aparece en los datos la dirección postal de los clientes, se evitarán los sesgos derivados de la ubicación geográfica.

En cuanto a la **transparencia**, se utiliza como modelo la regresión lineal, siendo éste un modelo tan simple como ampliamente conocido.