

Descripción del proyecto:

Para una tienda de ropa a medida, se desea aumentar las ventas evaluando donde concentrar los esfuerzos: mejorando la experiencia de los clientes al sitio web y a la aplicación móvil.

Descripción de las variables de entrada y de salida:

Variables de entrada (características):

Considero las siguientes variables de entrada (características) para el modelo:

1. **Identificador del cliente:** Un identificador único para cada cliente (x_1). Es evidente que este identificador debe ser de tipo “anónimo”, por ejemplo, podemos asociar un número consecutivo a cada cliente.
2. **Cantidad de dinero** que ha gastado cada cliente a la tienda (x_2).
3. **Tiempo dedicado** por cada cliente a interactuar con **la página web** (x_3).
4. **Tiempo dedicado** por cada cliente a interactuar con **la aplicación móvil** (x_4).
5. **Estado de membresía:** Información sobre si cada cliente es miembro o no, y posiblemente el tipo de membresía si hay diferentes niveles (x_5).

Otras variables de entrada que serían muy relevantes y podrían considerarse (de acuerdo con la definición del proyecto) son, por ejemplo: grado de aceptación de la aplicación web (x_6) y/o la aplicación móvil (x_7) por los clientes, grado de conformidad con el personal de la tienda con que interactúan (x_8), etc. Aunque en la definición de proyecto no dice explícitamente nada de que esos datos estén disponibles o no, puede que realmente no existan, por tanto, contaríamos solo con: x_1 , x_2 , x_3 , x_4 y x_5

Variable de salida (etiqueta):

1. Total de ventas de la tienda o por cliente.

Algoritmo sugerido:

El algoritmo que recomiendo es “**Regresión Ridge**” (que es una variante de la “**Regresión Lineal**”) por las siguientes razones:

- Esta bastante claro que las características (datos de entrada) tienen una relación continua y lineal con la variable de salida. Por ejemplo: a mayor tiempo dedicado por los clientes a la página web (x_3) o a la aplicación móvil (x_4) mayor deben ser las compras que hagan esos clientes y por tanto mayor será la venta de la tienda. Creo que, las personas en general no pierden tiempo en visitar sitios webs de tienda de ropa si no tienen la intención de comprar algo en ellas. Por tanto, un algoritmo del tipo de los de regresión lineal es lo ideal.
- Me parece que varias de las variables de entradas están altamente correlacionadas, por ejemplo, puede ser normal que los clientes consulten los mismos productos tanto por la página web como por la aplicación móvil antes de comprarlos finalmente. Esto se denomina multicolinealidad y el algoritmo “**Regresión Ridge**” es bueno para estas situaciones.

- En este caso también debemos tener en cuenta que puede haber clientes que hayan comprado muy pocas veces pero que se hayan gastado mucho dinero en esas pocas ocasiones. El algoritmo **“Regresión Ridge”** incluye un término de regularización que creo que es bueno para estos casos.

Aspectos claves a tener en cuenta para un uso ético y responsable de las técnicas de ML en este proyecto:

1. Los datos de entrada relativos a clientes no deben tener ningún valor que identifique unívocamente a cada cliente. Es decir, no deberían tener ninguno de los siguientes tipos de valores: DNI, nombre y/o apellidos, direcciones postales, direcciones de email, número de teléfono fijo/móvil, dirección IP, etc. A no ser que los clientes hayan consentido el uso de dichos datos personales. Este aspecto es el de la **privacidad**. Por consiguiente, solo podremos identificar a cada cliente de forma anónima por ejemplo con un número diferente para cada cliente.
2. No se deberían recopilar/procesar más datos de los necesarios para llevar a cabo el proyecto, pues eso puede conllevar al riesgo de recopilar algún dato personal de los clientes. También, los datos que usemos no deberían ser vendidos y/o traspasados a otras compañías y/o departamentos de la empresa que va a llevar a cabo el proyecto, sin el consentimiento de los usuarios. Este punto también tiene que ver con la privacidad de los datos.
3. Otro aspecto ético que es aplicable a este proyecto es el de la **“transparencia”**. Debemos poder explicar el funcionamiento del algoritmo y el modelo (sistema) para que sean comprensibles de manera que las decisiones tomadas por ese sistema puedan ser comprendidas y evaluadas por los usuarios y las partes afectadas.

A grandes rasgos solo veo como relevantes para este proyecto los puntos antes mencionados, pero en el caso de que se llevara a cabo el proyecto se podrían realizar evaluaciones y/o estudios más profundos acerca del impacto ético y social durante el desarrollo del sistema y antes de su implantación y también considerar las posibles consecuencias a largo plazo.