

Proyecto de ML para incrementar el gasto anual de los 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 asesorías 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. Tras la sesión de consultoría, los clientes pueden realizar pedidos de ropa a través de una aplicación móvil o del sitio web de la empresa.

Metodología Propuesta

El algoritmo más adecuado para este proyecto es un modelo de Regresión Lineal. Ya que la tarea implica predecir una variable continua (el gasto anual del cliente), ahora si los datos son simples y presentan una relación lineal entre las variables independientes (como características del cliente) y la variable objetivo (gasto anual), un modelo de regresión lineal puede ser suficiente.

Las métricas de evaluación que se buscarían mirar serían el **MAE (Mean Absolute Error)**, para medir el error promedio absoluto entre las predicciones y los valores reales, útil para interpretar el error en unidades monetarias. Otra a sería el **RMSE (Root Mean Squared Error)**, para penalizar más los errores grandes y obtener una visión general del rendimiento del modelo y el **R2 (Coeficiente de Determinación)**, para evaluar qué proporción de la variabilidad en el gasto anual es explicada por el modelo.

Interpretación de Resultados

Coeficientes del modelo de regresión lineal

Los coeficientes de un modelo de regresión lineal representan el cambio esperado en la variable objetivo (en este caso, el gasto anual del cliente, "Yearly Amount Spent") por cada unidad adicional de la característica correspondiente, manteniendo las demás variables constantes. Los Coeficientes calculados son:

Length of Membership: (61.4361.4361.43) Este coeficiente es **positivo** y el más grande en magnitud. Esto indica que, por cada año adicional que un cliente ha sido miembro, se espera que el gasto anual aumente en aproximadamente **\$61.43**, manteniendo las demás variables constantes. Los clientes más antiguos tienden a gastar más, lo cual sugiere que fortalecer la lealtad del cliente puede ser clave para aumentar los ingresos en el negocio.

Time on App: (39.2839.2839.28) Este coeficiente también es **positivo**, lo que significa que por cada hora adicional que un cliente pasa en la aplicación, se espera que su gasto anual aumente en aproximadamente **\$39.28**. Así que la interacción con la aplicación está asociada con mayores niveles de gasto. Esto podría indicar que la app es un canal de ventas importante y que mejorar la experiencia en la aplicación podría incrementar las compras.

Avg. Session Length: (24.7724.7724.77) Este coeficiente, aunque positivo, es el más pequeño en magnitud. Indica que, por cada minuto adicional que un cliente pasa en promedio por sesión, se espera un incremento de **\$24.77** en el gasto anual. Aunque la duración promedio de las sesiones está relacionada con un mayor gasto, su impacto es menor en comparación con el tiempo total en la app o la duración de la membresía.

Intercepto: (500.38500.38500.38) Representa el gasto anual promedio que un cliente realizaría si todas las características fueran iguales a **cero**. En el contexto del negocio, puede interpretarse como el gasto base que se espera, incluso sin influencia de las variables explicativas. Aunque es raro que todas las características sean exactamente cero, es un valor de referencia.

El signo positivo de los coeficientes sugiere que todas las características tienen una relación directa con el gasto anual. Esto significa que, a medida que cualquiera de estas características aumenta, también lo hace el gasto. Por otro lado, la magnitud en la **duración de la membresía** tiene el mayor impacto sobre el gasto anual, lo que sugiere que la lealtad del cliente es el factor más importante. Luego el **tiempo en la aplicación** también tiene un impacto significativo, pero menor que la membresía, mientras que la **duración promedio de las sesiones** tiene el menor impacto relativo.

Valor del MSE del modelo

En este caso, se obtuvo un **MSE = 162.52** significa que, en promedio, el error cuadrático entre las predicciones y los valores reales es relativamente bajo, considerando que el modelo logra un **R² de 0.9757**, lo que lo hace confiable para tomar decisiones empresariales, como segmentar clientes (identificar clientes de alto valor), predecir ingresos o planificar campañas de marketing personalizadas basadas en predicciones de gasto.

R² (Coeficiente de Determinación)

En este caso, el modelo tiene un **R² = 0.9757**, lo que significa que el 97.57% de la variabilidad en el gasto anual del cliente está explicado por las características del modelo. Solo el 2.43% de la variabilidad es atribuible a factores no capturados por el modelo. Este alto valor de R² indica que el modelo estaría muy bien ajustado a los datos.

La mayoría de los patrones en el gasto anual del cliente se explican correctamente usando las variables "Length of Membership", "Time on App" y "Avg. Session Length". Sin embargo, el R² alto no garantiza que el modelo sea perfecto. Es importante verificar que no haya sobreajuste (el modelo se ajusta demasiado a los datos de entrenamiento) o los errores sean aceptables en el conjunto de prueba (lo cual ya se evaluó con el MSE).

Conclusión:

El modelo es altamente efectivo para predecir el gasto anual de los clientes, proporcionando un ajuste sólido y métricas de error aceptables. Las acciones sugeridas para la tienda serían:

Retención: Enfocarse en mantener a los clientes por más tiempo, pues el impacto de la membresía es el más fuerte.

Optimización digital: Maximizar el tiempo que los clientes pasan en la app mediante mejoras en la experiencia del usuario.

Predicción: Usar el modelo para identificar clientes con mayor probabilidad de aumentar su gasto y dirigir campañas de marketing específicas.

Presentado por: Stephanie Rodriguez