



Módulo 5 – Aprendizaje de Máquina Supervisado

Actividad: Caso E-commerce

Especialización en Ciencia de Datos

Actividad: Caso E-commerce SKLearn



Caso Compañía E-commerce

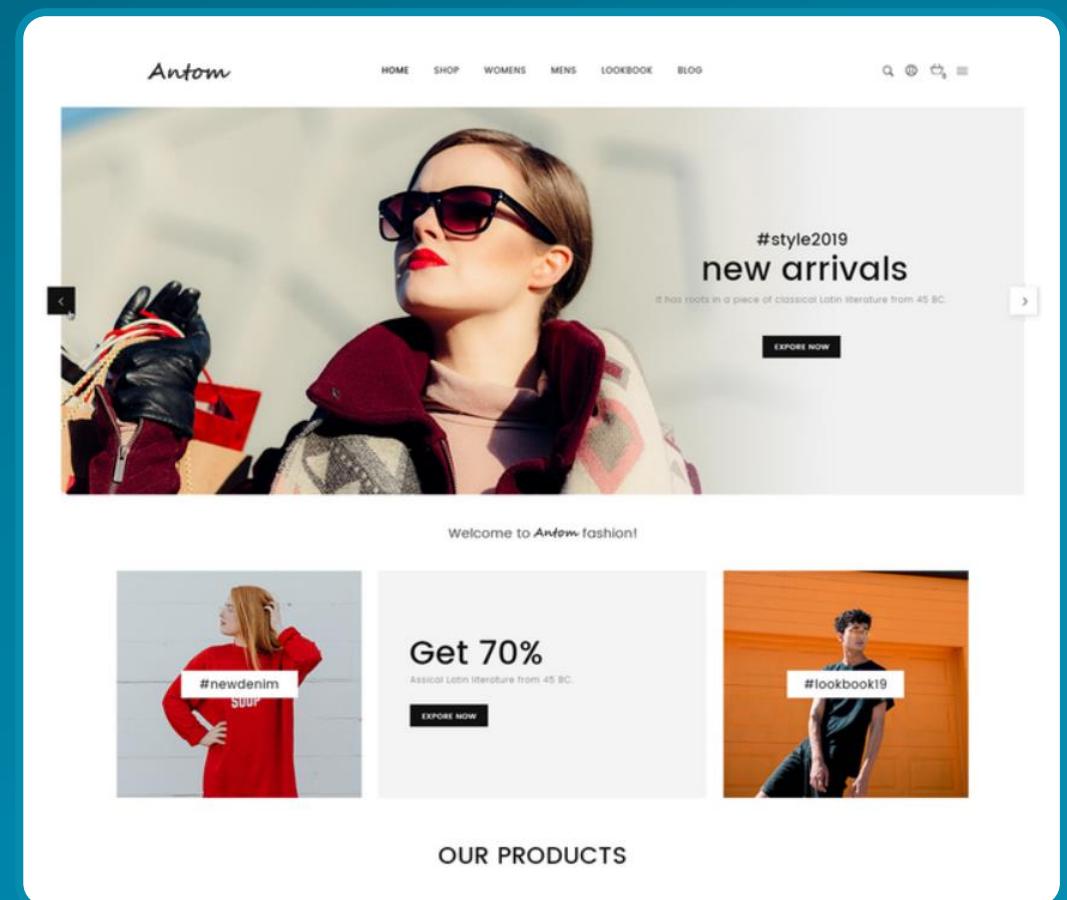
En esta actividad, tomaremos el set de datos de la compañía de E-commerce para elaborar un modelo regresivo, utilizando la librería scikit-learn y los conceptos aprendidos.

Email	Address	Avatar	Avg. Session Length	Time on App	Time on Website	Length of Membership	Yearly Amount Spent
mstephenson@fernandez.com	835 Frank Tunnel\nWrightmouth, MI 82180-9605	Violet	34.497268	12.655651	39.577668	4.082621	587.951054
hduke@hotmail.com	4547 Archer Common\nDiazchester, CA 06566-8576	DarkGreen	31.926272	11.109461	37.268959	2.664034	392.204933
pallen@yahoo.com	24645 Valerie Unions Suite 582\nCobbborough, D...	Bisque	33.000915	11.330278	37.110597	4.104543	487.547505
riverarebecca@gmail.com	1414 David Throughway\nPort Jason, OH 22070-1220	SaddleBrown	34.305557	13.717514	36.721283	3.120179	581.852344
mstephens@davidson-herman.com	14023 Rodriguez Passage\nPort Jacobville, PR 3...	MediumAquaMarine	33.330673	12.795189	37.536653	4.446308	599.406092

Caso Compañía E-commerce

Recordemos los features que tienen este set de datos:

- **Avg. Session Length:** tiempo promedio de las sesiones asesoramiento de estilo que se realizan en la tienda (minutos).
- **Time on App:** tiempo promedio de permanencia en la App mobile en minutos.
- **Time on Website:** tiempo promedio de permanencia en el sitio web en minutos.
- **Length of Membership:** tiempo de membresía, es decir, cuántos años el cliente ha sido miembro.
- **Yearly Amount Spent:** promedio de gasto en compras realizado de forma anual, en dólares.



Instrucciones

Usted deberá realizar lo siguiente:

1. Confeccione un notebook, ordenado y reproducible, en donde realizará un **modelamiento regresivo de aprendizaje de máquina** utilizando la librería SK Learn.
2. Determine los **pesos de cada feature** que componen el modelo.
3. A partir del punto anterior, proceda a la **selección de features**.
4. Realice el proceso de **validación cruzada** para determinar el nivel de ajuste del modelo y haga la comparación de resultados entre train y test set.
5. Determine las **métricas de error** tanto, train como en test.
6. Haga los **diagramas de error y de residuales** en el set de test.
7. En base a lo anterior, ¿**Se puede decir que el modelo es generalizable?**



25%

50%

15%

¡Éxito!