

Proyecto de ML para incrementar el gasto anual de clientes en la tienda.

Metodología Empleada.

En primer lugar, dividimos los datos en datos de entrenamiento y de prueba en una relación 80-20%. A continuación se comprueba la existencia de outliers y se winsorizan los campos 'Avg. Session Length', 'Time on App', 'Time on Website' y 'Length of Membership'.

Se imprime la matriz de correlación y se comprueba que 'Length of Membership', 'Time on App' y 'Avg. Session Length' guardan una correlación lineal con 'Yearly Amount spent'. Se seleccionan las tres primeras para entrenar un modelo de **regresión lineal**. Dichas variables son estandarizadas y se aplican todas las transformaciones a los datos de entrenamiento y de prueba.

Para evaluar nuestro modelo se utilizarán las métricas 'MSE' (error cuadrático medio) y R^2 (coeficiente de determinación).

Se separan las características de la variable objetivo y se aplica una evaluación cruzada con 10 'k-folds'. Luego se entrena el modelo con todos los datos de entrenamiento y se realiza una predicción. Ambos servirán para evaluar dicho modelo.

Evaluación del Modelo.

Validación cruzada.

MSE para cada conjunto: 128.3, 107.8, 137.8, 105.0, 91.2, 75.5, 149.6, 130.1, 91.7 y 103.8.

Media para todos los conjuntos: 112.1.

Desviación estándar para todos los conjuntos: 22,3

Se observa que los valores para el error cuadrático son relativamente pequeños y muestran poca dispersión.

R^2 Score de cada conjunto: 0.983, 0.981, 0.980, 0.974, 0.981, 0.990, 0.980, 0.979, 0.984, 0.980.

R^2 Score Promedio: 0.981

Como se ve, R^2 para todos los conjuntos de datos es muy próximo a 1.0.

Validación completa.

MSE: 162.5

R^2 Score: 0.976

Vuelve a comprobarse que el error MSE es pequeño y el coeficiente R^2 próximo a 1.

Coefficientes de regresión lineal: 61.43, 39.28, 24.77.

Intercept: 500.38.

Se muestra en el "notebook" la gráfica de los valores reales respecto de la predicción, que recalca la pequeña desviación de los errores y por tanto que el modelo está bien ajustado.

El modelo muestra una correlación lineal entre 'Length of Membership', 'Time on App' y 'Avg. Session Length' con 'Yearly Amount spent'. Por tanto, un mayor tiempo como usuario registrado en la aplicación WEB y en la aplicación móvil, aumentan la probabilidad de gasto en la tienda. Asimismo, dicha probabilidad también se ve aumentada con una mayor interacción del cliente con la WEB y la aplicación de móvil.