

Proyecto 1: Análisis y Predicción de Ventas en una Tienda de Retail

Proyecto 1- Parte final (Core)

Introducción

El dataset de ventas en diferentes sucursales contiene datos demográficos de clientes y ventas minoristas, este conjunto de datos permite identificar patrones, extraer información y obtener una comprensión del comportamiento del cliente.

Objetivo: El conjunto de datos ofrece un amplio espectro de factores que son de gran aplicabilidad en el mundo de las ventas.

La visualización de datos, el análisis estadístico y las correlaciones permite mostrar los matices que definen las operaciones minoristas y la dinámica de los clientes.

Información de los datos

Acerca de los datos:

ID de transacción: Identificador único para cada transacción que permite el seguimiento y la referenciación.

Fecha: En la que se realizó la transacción, esto indica información sobre las tendencias de ventas a lo largo del tiempo.

Identificación del cliente: Único para cada cliente, lo que contribuye un análisis orientado al cliente.

Género: Ya sea hombre o mujer, que ofrece información sobre patrones basados en el género.

Edad: Facilita la segmentación del cliente y la exploración de influencias relacionadas con la edad.

Categoría de Producto: Ayuda a comprender las preferencias de productos.

Cantidad: Número de unidades del producto compradas, permite tener información de volúmenes de compra.

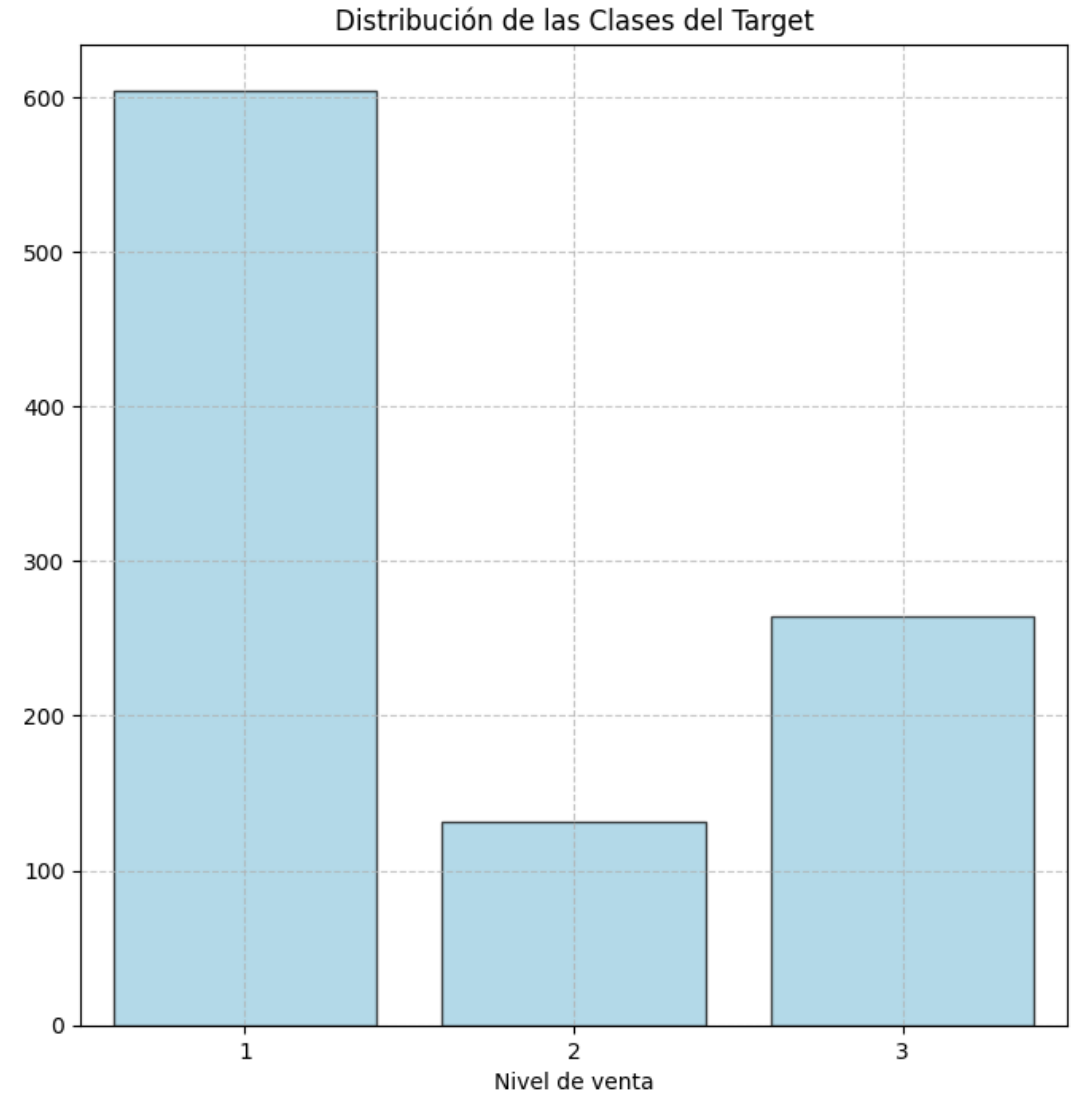
Precio por Unidad: Ayuda en los cálculos relacionados con el gasto total.

Monto Total: Total de la transacción que muestra el impacto financiero de cada compra.

Distribución de las clases 'Target'

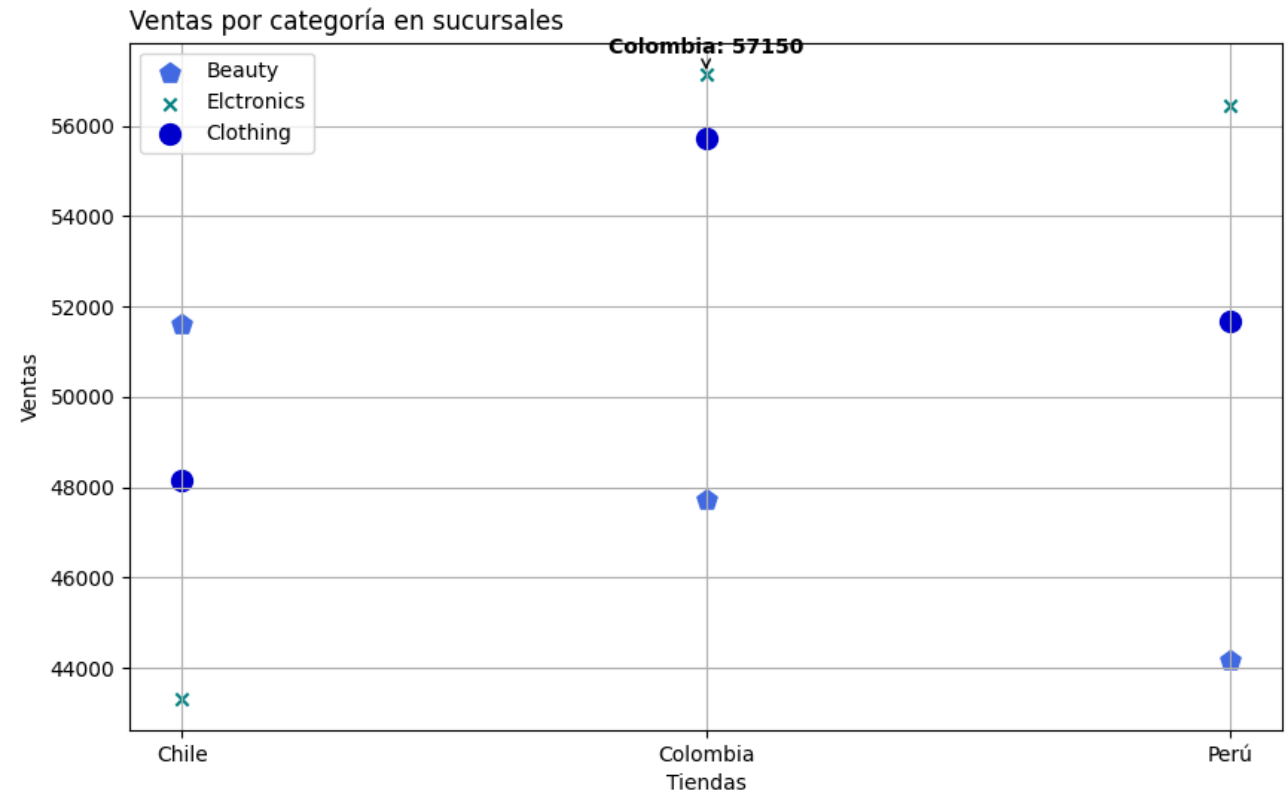
Se puede evidenciar una distribución bastante diferente de las categorías '1' para baja, '2' para media y '3' para venta alta.

Aunque se evidencia una distribución dispareja entre las clases, se trata de la clasificación del nivel de venta, lo cual el modelo conforme a los datos iniciales puede aprender de dicho comportamiento y lograr predecir clasificaciones futuras.



Presentación ventas sucursales

En el gráfico anterior se puede apreciar que la categoría 'Electronics' en la tienda 'Colombia' tiene mayor número de ventas al comparar con las tiendas 'Perú' y 'Chile'.



Modelo Random Forest Regressor

Mejores hiperparámetros: {'classifier__max_depth': 5, 'classifier__n_estimators': 50}

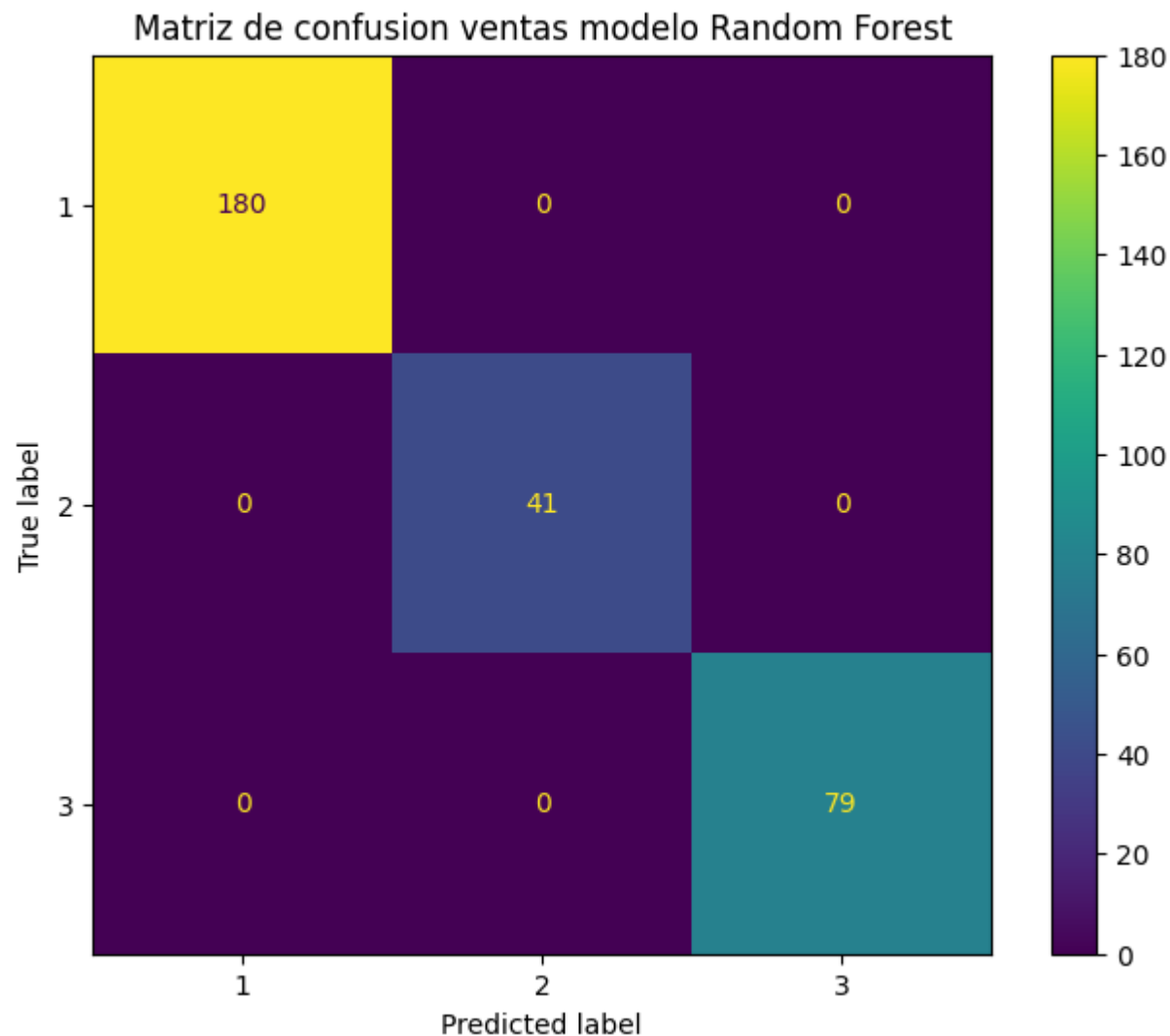
	precision	recall	f1-score	support
1	1.00	1.00	1.00	180
2	1.00	1.00	1.00	41
3	1.00	1.00	1.00	79
accuracy			1.00	300
macro avg	1.00	1.00	1.00	300
weighted avg	1.00	1.00	1.00	300

Una Exactitud: 1.0
Precisión, Recall y F1-Score: 1.00 para todas las clases.
Esto indica un ajuste perfecto a los datos, muestra un rendimiento impresionante, claro está que se ha realizado la validación cruzada para evaluar la robustez del modelo.

Gráfico Modelo Random Forest

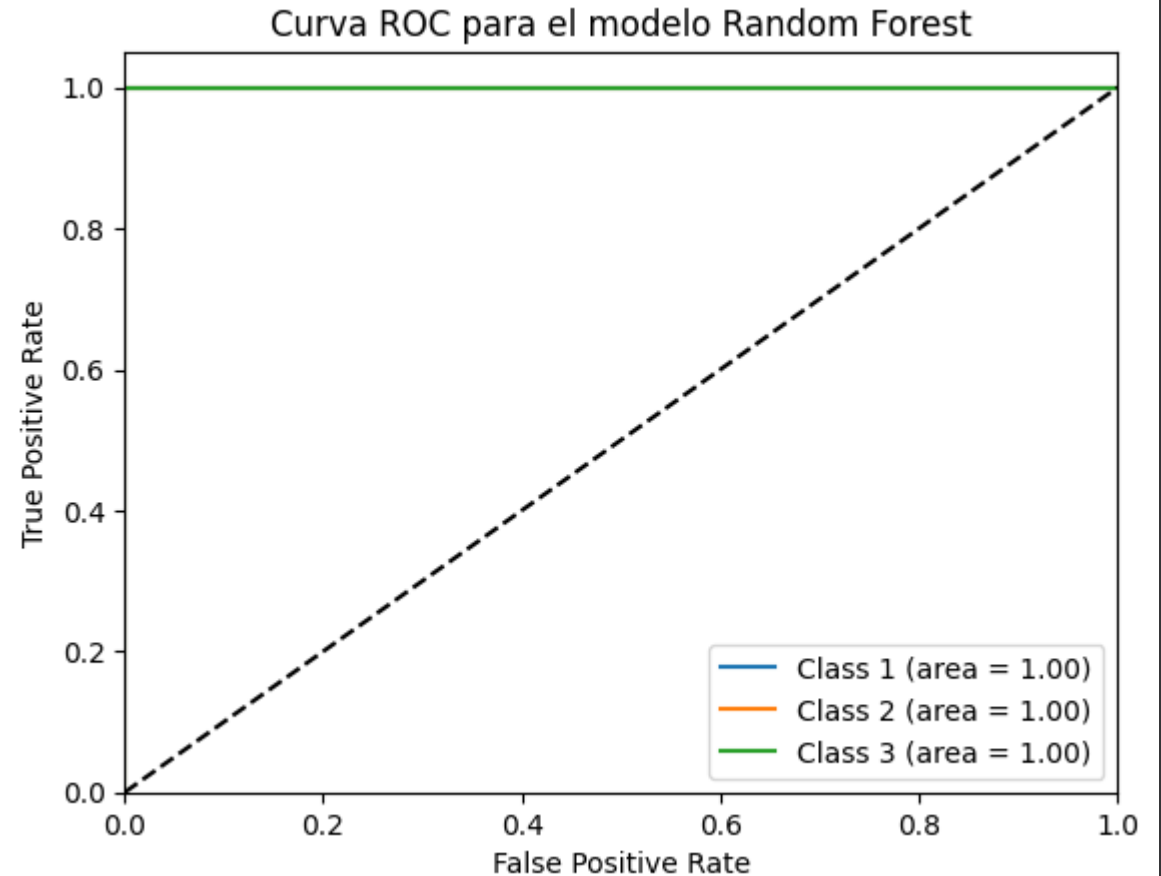
Perfecta Precisión y Recall, la matriz muestra que todas las instancias de cada clase fueron clasificadas correctamente.

La precisión y Recall para cada clase son 1.00 ya que no presenta errores de clasificación.



Curva ROC modelo Random Forest

Las curvas expresadas de la forma anterior son prueba de que el conjunto de datos es correcto, entonces es posible que el problema de clasificación sea sencillo y que el modelo logra clasificaciones perfectas.



Conclusión

Random Forest y Regresión Logística presentan un rendimiento perfecto, al realizar la verificación y validación, los resultados indican una robustez de los modelos, lo cual asegura que el rendimiento no se deba a la memorización o a un conjunto de prueba no representativo. En términos generales Random Forest y Regresión Logística muestran un rendimiento impresionante.

Sería interesante alimentar la data con más datos para realizar ajustes durante la marcha, en términos generales el modelo presenta un comportamiento aceptable dadas las circunstancias de la distribución de las clases, lo que no fue impedimento para que lograra adaptar de forma homogénea el modelamiento de la información.