



# PREDICCIÓN DE VENTAS PARA PRODUCTOS ALIMENTICIOS DE UNA CADENA TIENDAS

**Presenta: Mariana Arismendi**



B

J

E

T

I

V



S

## GENERAL

Desarrollar diferentes modelos de machine learning que permitan predecir el volumen de ventas para productos alimenticios que se venden en distintas tiendas

## ESPECÍFICOS



Presentar las características del dataset



Realizar una predicción de las ventas con diferentes modelos

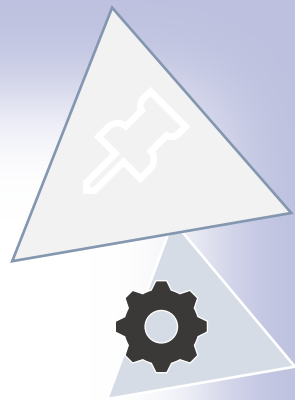
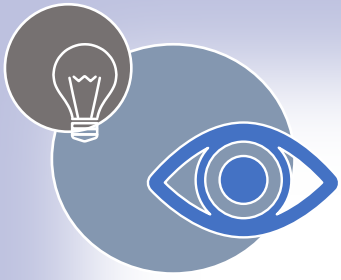


Ajustar los modelos



Evaluar los modelos y determinar el que mejor se ajusta





# DATOS DEL CONJUNTO DE DATOS

Nombre de la variable	Descripción
Item_Identifier	Identificación única del producto
Item_Weight	Peso del producto
Item_Fat_Content	Si el producto es bajo en grasa o regular
Item_Visibility	El porcentaje de la superficie total de exposición de todos los productos de una tienda asignada al producto concreto
Item_Type	La categoría a la que el producto pertenece
Item_MRP	Precio máximo de venta al público (precio de catálogo) del producto
Outlet_Identifier	Identificación única de la tienda
Outlet_Establishment_Year	El año cuando se estableció la tienda
Outlet_Size	El tamaño de la tienda con respecto a la superficie
Outlet_Location_Type	El tipo de área donde está ubicada la tienda
Outlet_Type	Si el mercado es un almacén o algún tipo de supermercado
Item_Outlet_Sales	Ventas del producto en la tienda particular Es la variable objetivo a predecir



# Ejecución de los Algoritmos

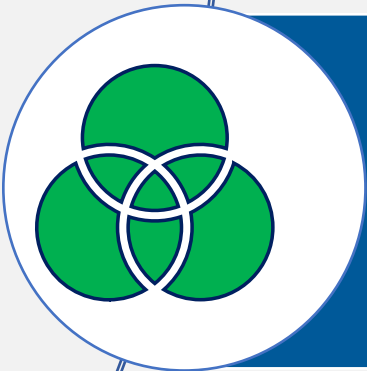


El tamaño del dataset a modelar después del preprocesamiento es de (6392, 39).

La semilla para ambos modelos se fijó en 42



Las métricas de regresión analizadas son las del error cuadrático medio



Para el modelo de árbol de regresión se realiza un gráfico que nos ayude a mostrar para qué profundidad tenemos un mejor  $R^2$  y así ajustar el modelo a este para hacerlo más óptimo.



# EVALUACIÓN COMPARATIVA DE LOS MODELOS



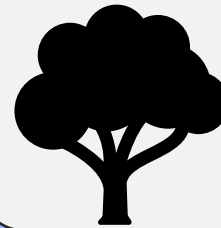
## • REGRESIÓN LINEAL



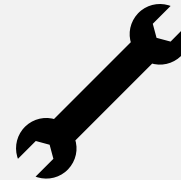
- BAJA VARIANZA DE LOS R2 :
- DATOS DEL TEST : 0.567
- DATOS DEL TRAIN : 0.561



- RMSE TRAIN : 1139.12
- RMSE TEST : 1092.86



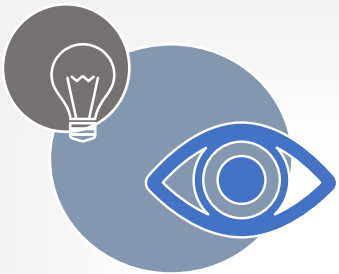
## • ÁRBOL DE REGRESIÓN



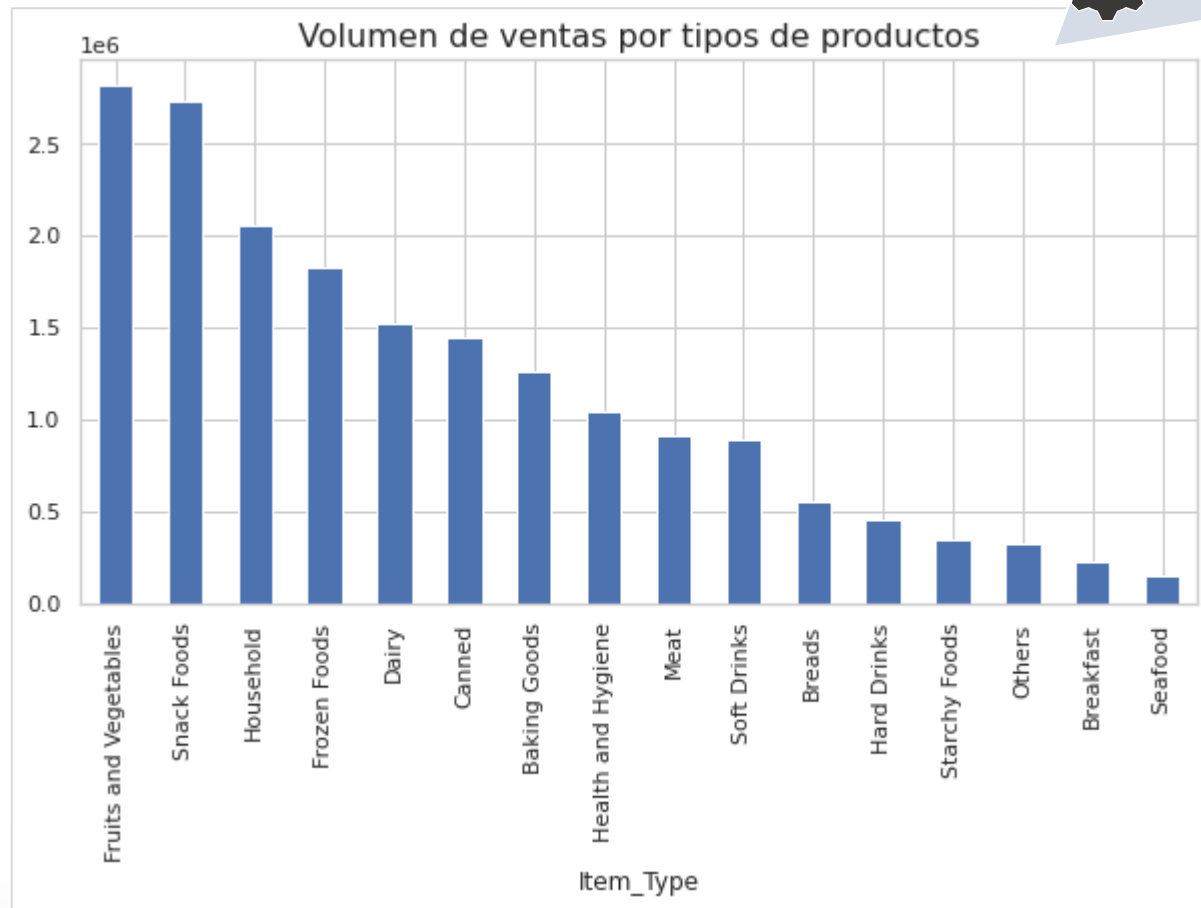
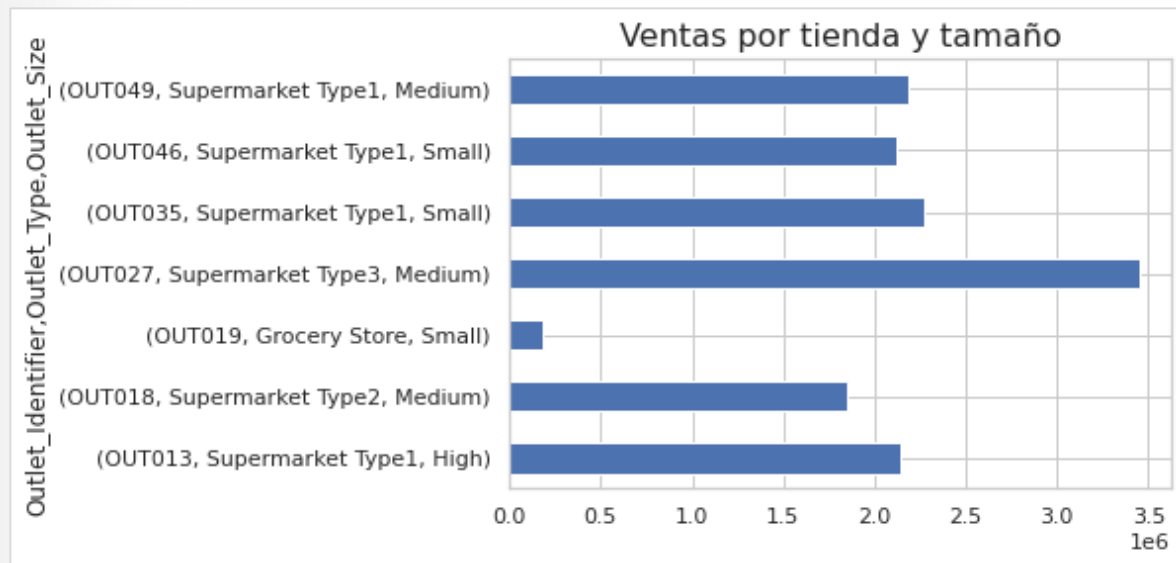
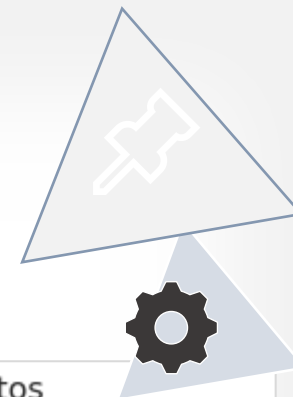
- MODELO SOBREAJUSTADO
- SE AJUSTA HALLANDO GRÁFICAMENTE EL MAX\_DEPTH PARA EL R2 ÓPTIMO :
- DATOS DEL TEST : 0.604
- DATOS DEL TRAIN : 0.595



- RMSE TRAIN : 4.93 E-15
- RMSE TEST : 1058.33



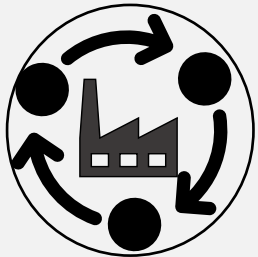
# DATOS GRÁFICOS RELEVANTES



# CONCLUSIONES



Árbol de  
regresión



MEJOR  
MODELO

R2 POCO  
EFICIENTE



ERROR  
CUADRÁTICO  
MEDIO MUY  
ALTO



Baja  
capacidad  
de predcción



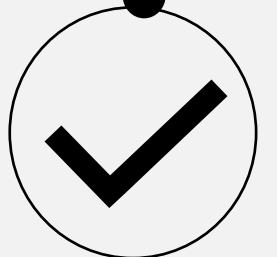
DECISIONES DE  
NEGOCIO  
SESGADAS

MUCHA  
ALEATORIEDAD  
EN LOS DATOS



LAS  
PREDICCIONES  
NO SON  
OBJETIVAS

Mejorar los  
datos para  
garantizar  
mejores  
resultados



Realizar  
análisis MÁS  
PROFUNDO y  
por segmentos