Big Mart Sales Prediction (Regression)

Introduction

The data scientists at BigMart have collected 2013 sales data for **1,559** products across **10** stores in different cities. Also, certain attributes of each product and store have been defined. The aim is **to build a predictive model and find out the sales of each product at a particular store** (**Item_Outlet_Sales**). Using this model, BigMart will try to understand the properties of products and stores which play a key role in increasing sales.

Please note that the **data may have missing values as some stores might not report all the data due to technical glitches**. Hence, it will be required to treat them accordingly.

Data

- **Item_Identifier** – Unique product ID, alpha string of length 5, 1559 distinct values;
- **Item_Weight** – Weight of the product;
- **Item_Fat_Content** – Categorical and describes whether the product is low fat or not:
  - Low Fat, LF
  - Regular, reg
- **Item_Visibility** – Percent of the display area in a store allocated to the product;
- **Item_Type** – Categorical and describes the food category to which the item belongs. 16 categories listed as follows: ['Dairy', 'Soft Drinks', 'Meat', 'Fruits and Vegetables', 'Household', 'Baking Goods', 'Snack Foods', 'Frozen Foods', 'Breakfast', 'Health and Hygiene', 'Hard Drinks', 'Canned', 'Breads', 'Starchy Foods', 'Others', 'Seafood'];
- **Item_MRP** – Maximum Retail Price (list price) of the product
- **Outlet_Identifier** – Unique store ID, alphanumeric string of length 6, 10 distinct values;
- **Outlet_Establishment_Year** – The year in which store was established;
- **Outlet_Size** – Size of the store; Categorical: ['High', 'Medium', 'Small'];
- **Outlet_Location_Type** – Categorical and tells the size of the city in which the store is located: ['Tier 1', 'Tier 2', 'Tier 3'];
- **Outlet_Type** – Categorical and tells whether the outlet is just a grocery store or some sort of supermarket. Following are the 4 categories in which the data is divided: ['Supermarket Type1', 'Supermarket Type2', 'Grocery Store','Supermarket Type3'];
- **Item_Outlet_Sales** – Outcome to be predicted; the sales of the product in the store.

Goal

Our Goal is to predict **Item_outlet_sales** with regression model of choice.

Data Source

https://www.kaggle.com/datasets/brijbhushannanda1979/bigmart-sales-data?resource=download