**BIG MART SALES PREDICTION**

**DISCRIPTION**

The aim is to build a predictive model and find out the sales of each product at a particular store. Create a model by which Big Mart can analyse and predict the outlet production sales.

A perfect project to learn Data Analytics and apply Machine Learning algorithm (XGBoost) to predict the outlet production sales.

**DATASET DISCRIPTION**

Big Mart has collected sales data from the year 2013, for 1559 products across 10 stores in different cities. Where the dataset consists of 12 attributes like Item Fat, ItemType, Item MRP, Outlet Type, Item Visibility, Item Weight, Outlet ldentifier, Outlet Size, Outlet Establishment Year, Outlet Location Type, Item ldentifier and Item Outlet Sales. Out of these attributes response variable is the Item Outlet Sales attribute and remaining attributes are used as the predictor variables.

The data-set is also based on hypotheses of store level and product level. Where store level involves attributes like city, population density, store capacity, location, etc and the product level hypotheses involves attributes like brand, advertisement, promotional offer, etc.

**DATASET**

<https://www.kaggle.com/datasets/shivan118/big-mart-sales-predictiondatasets?resource= download>

**DATASET DETAILS**

The data has 8523 rows of 12 variables.

**Variable Details**

* Item\_Identifier - Unique product ID
* Item\_Weight - Weight of product
* Item\_Fat\_Content - Whether the product is low fat or not
* Item\_Visibility - The % of the total display area of all products in a store allocated to the particular product
* Item\_Type - The category to which the product belongs
* Item\_MRP - Maximum Retail Price (list price) of the product
* Outlet\_Identifier - Unique store ID
* Outlet\_Establishment\_Year - The year in which store was established
* Outlet\_Size - The size of the store in terms of ground area covered
* Outlet\_Location\_Type -The type of city in which the store is located
* Outlet\_Type - whether the outlet is just a grocery store or some sort of supermarket
* Item\_Outlet\_Sales - sales of the product in t particular store. This is the outcome variable to be predicted.

**ADVANTAGES**

1. Optimized Inventory Management
2. Effective Marketing Strategies
3. Enhanced Store Performance
4. Waste Reduction
5. Business Planning and Budgeting

**PROJECT OVERFLOW**

1. Loading Packages
2. Loading and Analysing Data
3. Missing Value Treatment
4. Exploratory Data Analysis
5. Pre-processing Data
6. Need Of transformation – One Hot Encoding
7. XGBoost Regression Model
8. Creating Pipeline
9. Predicting on Test Data
10. Saving the Prediction in Excel Format