

# Case Study-Summary

## Project Overview

The objective was to predict sales for each product at Big Mart outlets using supervised machine learning techniques. The approach included thorough data cleaning, feature engineering, feature scaling, model training, hyperparameter tuning, evaluation, and final prediction submission.

### 1. Data Cleaning

- Loaded and verified datasets.
- No duplicates found.
- Handled missing values: Item\_Weight (interpolated), Outlet\_Size (mode-based filling), and Item\_Visibility (interpolated after replacing zeros).

### 2. Feature Engineering

- Fixed inconsistent categories in Item\_Fat\_Content.
- Created multiple new features: Outlet\_Type\_Size\_Location, Weight/Visibility, Outlet\_Age, Item\_Type\_Category, Category\_MRP\_Mean, Outlet\_Group\_MRP\_Mean, Outlet\_Visibility\_Mean, and Item\_Outlet\_Visibility\_Deviation.

### 3. Feature Encoding and Scaling

- Label Encoded all categorical features after combining train and test.
- Applied StandardScaler: fitted only on Train, then transformed both Train/Test.

### 4. Model Building and Training

- Trained models: Linear Regression, Ridge, Lasso, Decision Tree, Random Forest, XGBoost.
- Applied RandomizedSearchCV for hyperparameter tuning with 5-fold cross-validation.

### 5. Model Evaluation

- Used Root Mean Squared Error (RMSE) and R2 Score.
- Compared model performances.
- Random Forest achieved the best balance of low RMSE and high R2.

### 6. Final Prediction and Submission

- Best model (Random Forest) used for final prediction.
- Generated submission file with Item\_Identifier, Outlet\_Identifier, and Item\_Outlet\_Sales.

## **Conclusion**

- Thorough pipeline from data preparation to model evaluation.
- Strong feature engineering and clean modeling approaches.