

BigMart Sales Prediction: Approach & Methodology

Project Overview

Our goal is to forecast sales for BigMart outlets using data-driven machine learning methods. This initiative seeks to aid business decisions by understanding what drives sales and providing accurate predictions for individual outlet-item combinations.

Key Challenges

- **Missing Data:** Some features, notably Item_Weight (~17%) and Outlet_Size (~28%), have substantial gaps.
- **Data Quality:** Inconsistencies include zero values for Item_Visibility and variations in how Item_Fat_Content is recorded.
- **Feature Variety:** The dataset mixes numeric and categorical features.

Data Cleaning & Feature Engineering

- **Filling in Gaps:** For missing Item_Weight values, we use averages grouped by Item_Identifier. Outlet_Size gaps are addressed based on similar outlets.
- **Visibility Fixes:** Zeros in Item_Visibility are replaced with average visibility for that specific item.
- **Categorical Consolidation:** To improve consistency, Item_Fat_Content categories are standardized (e.g., 'Low Fat' & 'LF' merged), reducing noise.
- **Food/Non-Food Split:** We combine item types into broader 'Food' and 'Non-Food' categories, helping the model distinguish product classes more effectively.

Modeling Approach

After thorough data prepping:

- **Exploratory Analysis:** We examine feature distributions and correlations, noting the target (Item_Outlet_Sales) is strongly right-skewed.
- **Model Selection:** Multiple algorithms are tested, beginning with linear regression, progressing to robust tree-based models (e.g., Random Forest, XGBoost) to handle non-linearity and interactions.
- **Validation:** Performance is evaluated via cross-validation, aiming to minimize prediction error and avoid overfitting.

Summary

By carefully cleaning data, engineering meaningful features, and leveraging powerful models, we build a solution that addresses BigMart's sales prediction needs realistically and reliably. This approach combines business intuition with technical expertise, ensuring practical, actionable results for stakeholders.