Project Documentation

Project Title: Big Mart Sales prediction

Project Date: 12/02/2022

1. Code Documentation:

A detailed Explanation is given in the code cell please have a look.

4. Dataset and Results:

Please find the

1. **Dataset** is named as a **resources Folder** in the Zip file.

A detailed description of the Dataset you can find in the Presentation Slides.

1. **Result Datasets** is named as an **Output Folder** in the Zip file.

**Project Report:**

**project summary and Problem Statement:**

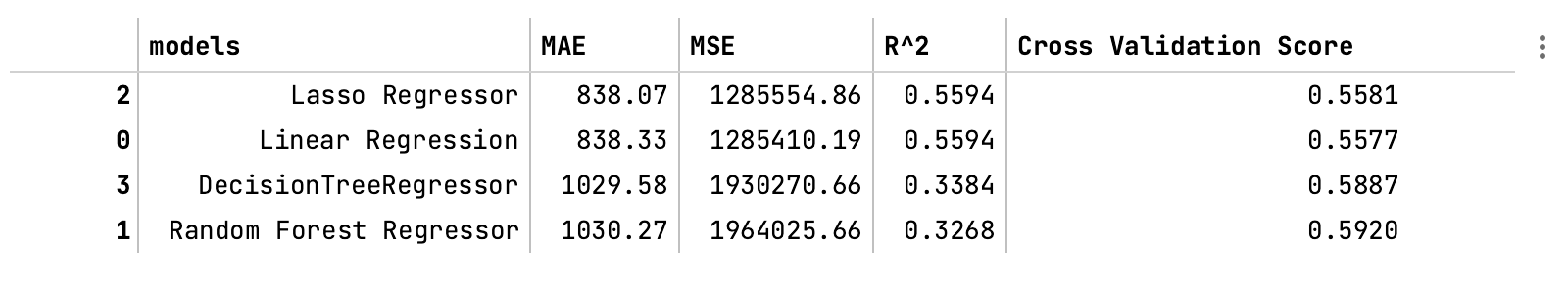
This paper analyses the Big Mart Sales Prediction, in an effort to create an effective sales predictor that can be used for business decision-making purposes

* The data scientists at BigMart have collected 2013 sales data for 1559 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. Using this model, BigMart will try to understand the properties of products and stores which play a key role in increasing sales

Realizations During Data Exploration:

* The difference in item types by sales is very small.
* Outlet 27 is the most profitable and there is a big difference between each specific outlet's sales.
* Surprisingly supermarket type 3 is the most profitable and not type 1.
* Medium and high outlet sizes are pretty much even in sales.
* Tier 2 and 3 are almost even the highest in sales (2 is slightly larger).

Results:

****

Realizations:

1. `Item\_MRP` optimizes Maximum Outlet sales (positive correlation with the target).
2. Linear Regression and Lasso Regressor have the best performance in most categories.
3. Only a third of the observed variation can be explained by the model's inputs of Random Forest Regressor, there for its performance is not optimal even though its cross-validation is the highest.
4. For better performance these models need tuning e.g. Grid Search.