Big Mart Sales Prediction

Online 26-05-2016 12:01 AM to 31-08-2020 11:59 PM

ENDS IN
31 DAYS 6 HOURS 29 MIN

37243 Registered

Practice Problem Prizes

Registered

Sales Prediction for Big Mart Outlets

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 predict the sales of each product at a particular outlet.

Using this model, BigMart will try to understand the properties of products and outlets 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 Dictionary

We have train (8523) and test (5681) data set, train data set has both input and output variable(s). You need to predict the sales for test data set.

Train file: CSV containing the item outlet information with sales value

Variable	Description
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 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 the particular store. This is the outcome variable to be predicted.

Test file: CSV containing item outlet combinations for which sales need to be forecasted

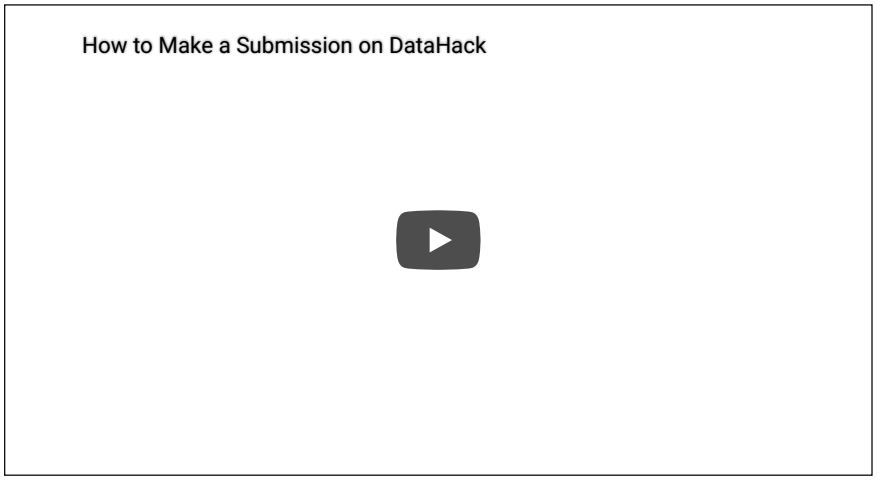
Variable	Description
Item_Identifier	Unique product ID

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 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

Submission file format

Variable	**Description**
Item_Identifier	Unique product ID
Outlet_Identifier	Unique store ID
Item_Outlet_Sales	Sales of the product in the particular store. This is the outcome variable to be predicted.

How to Make a Submission?



Evaluation Metric

Your model performance will be evaluated on the basis of your prediction of the sales for the test data (test.csv), which contains similar data-points as train except for the sales to be predicted. Your submission needs to be in the format as shown in sample submission.

We at our end, have the actual sales for the test dataset, against which your predictions will be evaluated. We will use the **Root Mean Square Error** value to judge your response.

Public and Private Split

Test file is further divided into Public (25%) and Private (75%)

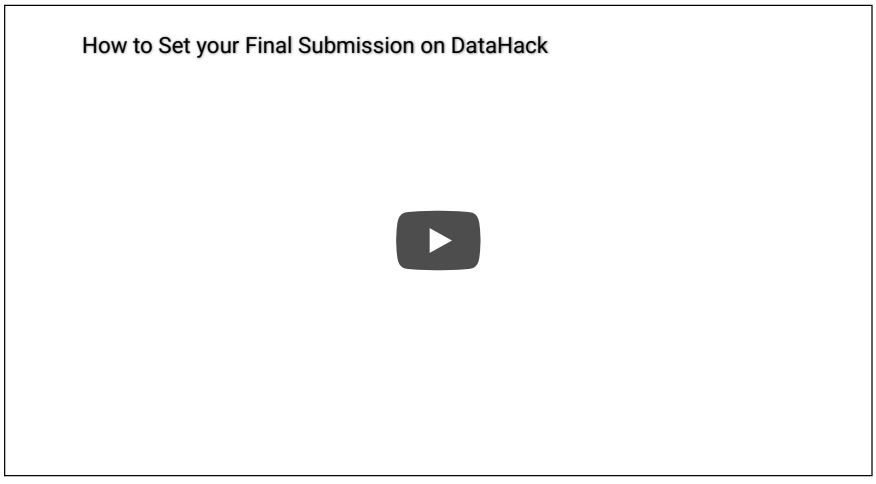
- Your initial responses will be checked and scored on the Public data.
- The final rankings would be based on your private score which will be published once the competition is over.

Guidelines for Final Submission

Please ensure that your final submission includes the following:

1. Solution file containing the predicted sales value in the test dataset (format is given in sample submission csv)
2. Code file for reproducing the submission, note that it is mandatory to submit your code for a valid final submission

How to Set Final Submission?



Data

Test File | Train File | Sample Submissions