Contents

I. Problem Description

Predicting the sales of products across stores of a retail chain

A large Indian retail chain has stores across 3 states in India: Maharashtra, Telangana and Kerala. These stores stock products across various categories such as FMCG (fast moving consumer goods), eatables / perishables and others. Managing the inventory is crucial for the revenue stream of the retail chain. Meeting the demand is important to not lose potential revenue, while at the same time stocking excessive products could lead to losses.

In this hackathon you are tasked with building a machine learning model to predict the sales of products across stores for one month. These models can then be used to power the recommendations for the inventory management software at these stores.

II. The datasets are provided as cited below:

Target attribute: "sales" (continuous)

train_data.csv:

- date : The date for which the observation was recorded
- product_identifier : The id for a product
- department_identifier : The id for a specific department in a store
- category_of_product : The category to which a product belongs
- outlet : The id for a store
- state : The name of the state
- sales : The number of sales for the product

Auxiliary Datasets:

- product_prices.csv : The prices of products at each store for each week
- date_to_week_id_map.csv : The mapping from a date to the week_id
- sample_submission.csv : The format for submissions

- The test_data.csv file has all the attributes of the train_data.csv file excluding the sales (target) column

III. Tasks:

Model Building:

You are expected to create an analytical and forecasting framework to predict the sales of the products based on the quantitative and qualitative features provided in the datasets. You may derive new features from the existing features and also from the domain knowledge, which may help in improving the model efficiency.

Visualization Tasks:

Exploratory Data Analysis using visualizations in R Notebook or Jupyter notebook format. (all train data to be used for this task)

- List down the insights/patterns observed from the visualizations
- Explain the impact of the most important attributes on the target attribute observed from data visualizations.

Observations:

Are there any overfitting or underfitting problems? If yes, how do you address it?

IV. Evaluation Metric:

- The evaluation metric for this hackathon is the RMSE or the Root Mean Squared Error.