INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

Walmart
Save money. Live better.

talent sprint

Session 15 Lab Abstract

Experiment 0:

Objective: As an AI / ML expert help a leading retailer forecast the sales.

Link: https://www.kaggle.com/t/364a1a5fb7ca4b0da0d3113fc9b6e639

Deadline: Thursday 9 pm, 10/May/2018 IST (with prize money!)

Forecasting is an important approach to plan the future effectively and efficiently. A leading retailer in USA, wants to forecast sales for their product categories in their store based on the sales history of each category. Sales forecast has very high influence on the performance of the company's business and hence these sales forecasts can be used to estimate company's success or performance in the coming year. Accurate forecasts may lead to better decisions in business. Sales or revenues forecasting is very important for retail operations. Forecasting of retail sales helps retailer to take necessary measures to plan their budgets or investments in a period (monthly, yearly) among different product categories like women clothing, men clothing and other clothing and at the same time they can plan to minimize revenue loss from unavailability of products by investing accordingly.

Dataset Release Schedule:

The train, test, and attribute description datasets have been released fully. In addition the below 3 datasets (Holiday, Weather, Macro-economic) have their heads released.

The additional full datasets for the same are released on following days and on a particular time (**VIA EMAIL ONLY**):

S.No.	Event	Time
1	Introduction to the Case Study, familiarity with the dataset, missing values, data preprocessing	Saturday 3 pm
2	Release Macro_economic_data	Sunday at 9:00 am
3	Release Weather_data	Monday at 1:00 pm
4	Release Holiday_data	Tuesday at 1:00 pm

Attributes:

1. Train.csv - This files contains the following attributes :
   1. Year
   2. Month
   3. Product Category
   4. Sales (In Thousand Dollars)
2. Event_Holiday.csv - This file contains the following attributes:
   1.Year
   2. Month Date
   3. Event
   4. Day Category
3. Weather_Data.csv - This file contains the following attributes :
   1. Year
   2. Day
   3. Month
   4. Temp high (°C)
   5. Temp avg (°C).... e.t.c.
4. Macroeconomic_Data.csv - This file contains the following attributes:
   1. Year-Month
   2. Monthly Nominal GDP Index (in Million$)
   3. Monthly Real GDP Index (in Million$) ….e.t.c.

Prizes: This competition has prize money and the rewards for the competition are written on the Kaggle website. This competition should be done in groups (groups that were already made before, individual submissions will not be considered). Final discretion of the leaderboard is in the hands of Talentsprint (honor code violation, etc). The prizes and certificates are given on the final day. This data is proprietary please do not share the dataset with anyone. The Solution Test file and the solution Python notebook will not be provided.

Experiment 1: (Ungraded & No submission at this time)

Please see the following link:
https://drive.google.com/file/d/1yZkfgtm_nhRW89zESqvHJ1DEuWNZWVQo/view?usp=sharing

Narrative: Over the week you will be given two independent tasks as preparatory steps towards the final Hackathon. While every one of you should try the first step, as long as one of members in each team is able to run the platform on their laptops successfully, you should be able to complete this mini-Hackathon. 1. Install these software platforms: Gazebo Simulator and ROS, you will be given further instructions on how to install these on your laptop as well as the option of running them from a bootable USB with a linux installation. Train a classifier to recognize a set of road signs from their images. The training images will be provided to you with the corresponding labels, and you are expected to develop a classifier that can label samples in a test set.

Good luck! Jay