Predicting the consumer’s next purchase order

Instacart Solution Report

Serhan Sari



# Problem Definition

Predict the user’s next purchase product(s) by analyzing for each user’s historical purchases

* Key Questions for solving the problem:
  + Are there any repetitive products by user? Is there any relation with day/time?
  + Is there any correlation between the purchased product(s) and the ordered day/time?
  + Is there any relation between aisles-departments and products?
  + What are the most and least frequent products which each user buy?
  + What is the time interval between user’s orders?
  + What is the time interval between two orders for same product?

# Dataset

There are 6 different files that consist of a relation between each other which describe consumers’ orders over time. There are more than 3 million grocery orders and from more than 200,000 anonymized Instacart users which contains each users’ past purchased products from 4 to 100 in each order by weekly, daily and hourly that the order was placed and also the interval time between the orders.

1. **aisles.csv**

* Aisle numbers and description;

Headers: aisle\_id, aisle

1. **deparments.csv**

* Department numbers and description;

Headers: department\_id, department

1. **order\_products\_prior.csv & order\_products\_train.csv**

* These files provide us in each order which product placed an order along reordered rate;

Headers: order\_id, product\_id, add\_to\_cart, reordered

1. **orders.csv**

* This file contains each and every user’s order history along with the order times (hour & day), reorder time interval between the last time ordered and which set (prior, train) an order belongs.

Headers: order\_id, user\_id, eval\_set, order\_number, order\_dow, order\_hour\_of\_day, days\_since\_prior\_order

1. **products.csv**

* This file describes which products belong to which aisle and department

Headers: product\_id, product\_name, aisle\_id, department\_id

Customer Classification? What percentage of the customer how frequently order and frequency of repetitiveness?

50 factors – key questions – anything which impacts

probability of user purchase?

Purchase probability by user?

Probability of reordering?