****

**Team 5 Final Project:**

**Instacart Market Basket Analysis**

11.23.2017

Video URL: https://www.youtube.com/watch?v=mqcd3F2N5Io&feature=youtu.be

Jaini Bhansali

Tushar Goel

Final Project: INFO 7390 Advances In Data Science

Overview

In the era of e-commerce where we buy anything we wished for in a span of a click, groceries is the new product that can be bought with a mouse click. Companies like Amazon Fresh, InstaCart are utilizing this to deliver groceries at the user’s doorstep.  Instacart, a grocery ordering and delivery app, aims to make it easy to fill your refrigerator and pantry with your personal favorites and staples when you need them. After selecting products through the Instacart app, personal shoppers review their order and do the in-store shopping and delivery for you.

This project would help us contribute to revolutionary concept of grocery shopping as it has been predicted that by the year 2025, grocery sales is going grab 20% of the market sales.

Goals

Our goal is to implement a recommendation system. This will allow Instacart to provide the right products and the right combination of products to the customer among the variety of products that are present. With the overload of information on the internet there are a variety of products to choose amongst and hence, a recommendation system makes it slightly easier for a customer to buy products online. Our goals are as follows:

1. To provide recommendation of Instacart products to a new customer
2. To provide recommendation of Instacart products to an existing customer

We also attempt to improve the usability of the website by implementing dashboarding.By achieving the above, Instacart provides a better shopping experience to the customer by providing recommendations , when he logs into the InstaCart App. Instacart also capitalizes on the recommendation as customer might buy the recommended combination of products together as well.

**Use Cases**

**Instacart Customer**

1. Improve the shopping experience of a new user in order to make them returning customers
2. This will also help keep existing user by providing recommendations when they return to Instacart

**Instacart**

1. Help improve a customer’s shopping experience
2. Improve sales with recommendations as a customers might purchase a product that is recommended

**Data**

1. **InstaCart Data :** <https://www.kaggle.com/c/instacart-market-basket-analysis/data>

The data is a relational set of files describing customers' orders over time. The goal of the competition is to predict which products will be in a user's next order. The dataset is anonymized and contains a sample of over 3 million grocery orders from more than 200,000 Instacart users. For each user, we provide between 4 and 100 of their orders, with the sequence of products purchased in each order. We also provide the week and hour of day the order was placed, and a relative measure of time between orders.

The files presents are as follows:

1. aisles.csv
2. departments.csv
3. order\_products.csv
4. orders.csv
5. products.csv
6. order\_train.csv
7. order\_prior.csv

Process Outline

1. Data Preprocessing

* Data Cleaning and Handle Missing Value Analysis
* Join the different csv’s to form a joint dataset

1. Exploratory Data Analysis
2. Recommendation Systems (User based Recommendations)
3. Design a Data Pipeline and a feasible system to implement this approach
4. Deploy the Model using Azure/AWS or another feasible approach
5. Build a web application to demonstrate prediction and recommendations

Recommendations will be provided on the following basis:

Basic Recommendation will be provided on the following basis:

1. Most Bought Product of all time
2. Most visited department
3. Most frequently bought products by the specific user
4. Most frequently bought products on that hour
5. Most frequently bought product on that day
6. Most frequently bought product on that hour of that day
7. Most Reordered product by that user

Models Used

1. Collaborative Filtering (Metrics used : Precision,Recall)
2. Apriori Algorithm (Metrics used : Support, Confidence and Lift)

If time permits we will try to implement order prediction as a nice to have feature.

# Milestones

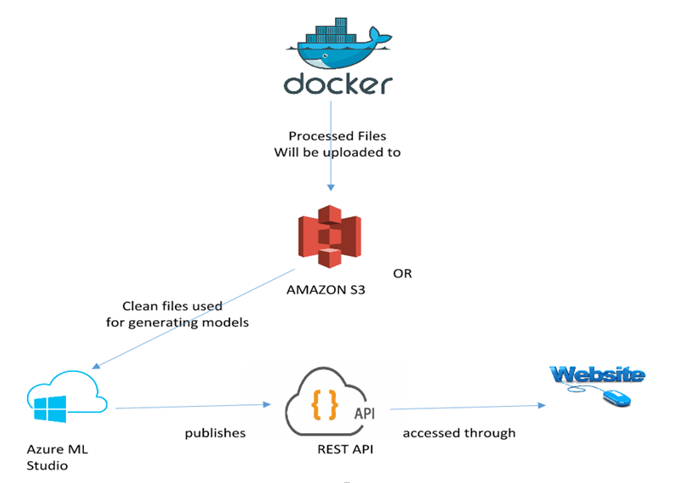
|  |  |
| --- | --- |
| **Timeframe** | **Delivery** |
| Day 1-4 | Data Preprocessing and Exploratory Data Analysis |
| Day 4-10 | Collaborative Filtering, Apriori (Association Rule Mining) |
| Day 10-15 | Deployment of models and Building Web Application |
| Day 15-20 | System Integration and Documentation |

Personas

1. Instacart Users : Will be presented with recommendations to improve their shopping experience
2. Instacart Admins: Dashboarding with summary metrics of sales

Deployment Details:

1. Language : Python, Java
2. Pipeline : LUIGI
3. Container Docker
4. Cloud Tools and Platform : Amazon S3, Microsoft Azure Machine Learning Studio, Aws EC2(yet to be completely finalized)
5. Tools for Analysis : Microsoft Azure Machine Learning Studio
6. Other considerations : Use of NodeJS



User Interface and Design Plan

Admin’s interface mockup



User Interface Mockup





References:

1. <https://www.kaggle.com/c/instacart-market-basket-analysis/data>
2. <https://www.instacart.com/datasets/grocery-shopping-2017>
3. <https://tech.instacart.com/3-million-instacart-orders-open-sourced-d40d29ead6f2>