****

**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

Instacart’s data science team plays a big part in providing this delightful shopping experience. Currently they use transactional data to develop models that predict which products a user will buy again, try for the first time, or add to their cart next grocery shopping experience

1. Our goal is to study orders over time to predict which previously purchased products will be in a user’s next order.
2. Also, we will predict if he would purchase the previous item again or also venture in to buy a new product.

By achieving the above, Instacart learns its stocking capacity for a particular season, and also in the future recommend these items to the customer, when he logs into the InstaCart App. This would be beneficial to the customer as well, as he would receive timely recommendations to improve his shopping experience.

**Use Cases**

Improving Instacart’s ability to predict a customers buy and in turn stock up accordingly with less wastage. Improving a User’s experience to shop at InstaCart.

**InstaCart:** Instatcart’s ability to predict a customers buy and in turn stock up accordingly in order to reduce wastage as these are perishable products

**InstaCart User**: Improve a user’s shopping experience with recommendations based on the buying pattern.

**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

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. Study of Supervised approaches
3. Time Series Analysis
4. Study of Unsupervised approaches
5. Design a Data Pipeline and a feasible system to implement this approach
6. Deploy the Model using Azure/AWS or another feasible approach
7. Build a web application to demonstrate prediction and recommendations

# Milestones

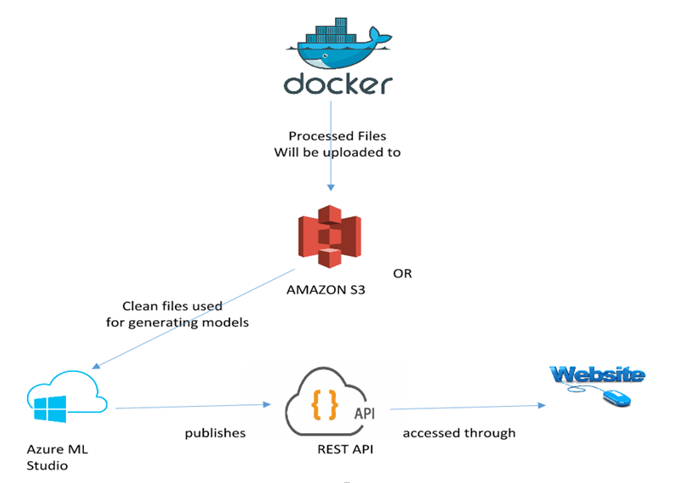
|  |  |
| --- | --- |
| **Timeframe** | **Delivery** |
| Day 1-4 | Data Preprocessing and Exploratory Data Analysis |
| Day 4-5 | Time Series Analysis |
| Day 5-10 | Training and Prediction using Supervised Learning, Time Series Analysis and Unsupervised Learning |
| Day 10-15 | Deployment of models and Building Web Application |
| Day 15-20 | System Integration and Documentation |

Personas

1. Instacart Employees : Employees will presented with the buying pattern of a user and hence, can manage and plan stocks for a season to reduce losses of perishable goods
2. Instacart Users : Will be presented with recommendations to improve their shopping experience

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





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>