Group member name: Hongqi.Guo // Jihui.Sheng

1) Data Mining Task: What is your data mining task?

Instacart Market Basket Analysis

This task could be a series of exploratory questions that you want to investigate or analyze. What is your motivation behind choosing this problem for your project?

we want to know what combinations of a product appear together in the entire supermarket , what combination it appears in a product together, and what combination it appears in an aisles.

For example: if we buy a bag of flour in the supermarket, then KitchenAid and whisk are appearing together in the whole supermarket.
Low flour flour and bread flour appear together in products.
Yeast and chocolate appear together in flour in aisles.

2) Dataset: What is the source of your data?

The Instacart Online Grocery Shopping Dataset 2017

This anonymized dataset contains a sample of over 3 million grocery orders from more than 200,000 Instacart users.

a) aisles.csv : Aisles ID that shows the aisle to which the product belongs
b) departments.csv : The category to which an item belongs.
c) products.csv: Product name. Is a collection of **aisles.csv** and **department.csv**.
d) order_products__prior.csv : Statistics of product purchases
e) orders.csv : Product purchase frequency recommendation
f) products.csv : Product ID and product category statistics

3) Methodology: How will you solve the data mining task?

You should have some idea of the algorithms or software tools you plan to investigate.

Please feel free to use existing data mining and machine learning tool kits (e.g., Weka, Scikit-Learn) as needed for your project.

We are going to use the Aprori algorithm for commodity correlation analysis.

4) Final product: What will be the outcome of this project?

The result of this project is a list of product recommendations after the user purchases an item.

Get a list of recommended products across the board.

Get a recommended list of products in the product.

Get a recommended list of products in aisles.

How will you measure the success of your course project?

Able to get accurate recommendation list.

Will this project help you explore or learn something new?

We can learn that using apriori algorithm must perform a lot of product grooming work, distinguish the hierarchical relationships of different levels, and label the corresponding products appropriately.