Assignment 11

1. Demonstrate the task of association rule mining on “The Instacart Online Grocery Shopping Dataset 2017”, accessed from <https://www.instacart.com/datasets/grocery-shopping-2017> dataset using apriori algorithm in python. Find out the support, Confidence and Lift parameters.

Problem Description: The dataset is a relational set of files describing customers' orders over time. The goal of this 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.

File descriptions

Each entity (customer, product, order, aisle, etc.) has an associated unique id. Most of the files and variable names should be self-explanatory.

aisles.csv

aisle\_id,aisle

1,prepared soups salads

2,specialty cheeses

3,energy granola bars

...

departments.csv

department\_id,department

1,frozen

2,other

3,bakery

...

orderproducts\_\*.csv

These files specify which products were purchased in each order. orderproductsprior.csv contains previous order contents for all customers. 'reordered' indicates that the customer has a previous order that contains the product. Note that some orders will have no reordered items. You may predict an explicit 'None' value for orders with no reordered items. See the evaluation page for full details. orderid,productid,addtocartorder,reordered

1,49302,1,1

1,11109,2,1

1,10246,3,0

…

orders.csv

This file tells to which set (prior, train, test) an order belongs. You are predicting reordered items only for the test set orders. 'orderdow' is the day of week. orderid,userid,evalset,ordernumber,orderdow,orderhourofday,dayssincepriororder

2539329,1,prior,1,2,08,

2398795,1,prior,2,3,07,15.0

473747,1,prior,3,3,12,21.0

…

products.csv

product\_id,product\_name,aisle\_id,department\_id

1,Chocolate Sandwich Cookies,61,19

2,All-Seasons Salt,104,13

3,Robust Golden Unsweetened Oolong Tea,94,7

...

sample\_submission.csv

order\_id,products

17,39276

34,39276

137,39276

1. Implementation of association rule using apriori algorithm in python on ‘Adult Census Income data’ dataset accessed from <https://www.kaggle.com/uciml/adult-census-income> and verify the result with Weka Tool.