

**ASSIGNMENT SUBMISSION FORM**

***This will be the first page of your assignment***

Course Name:                       Machine Learning – Unsupervised Learning 2

Assignment Title:                  Individual Assignment

Submitted by: Jhanvi Sharma

(Student name or group names)

|  |  |
| --- | --- |
| **Student Name** | **PG ID** |
| Jhanvi Sharma | 12120086 |
|  |  |
|  |  |
|  |  |

**ISB Honour Code**

* I will represent myself in a truthful manner.
* I will not fabricate or plagiarise any information with regard to the curriculum.
* I will not seek, receive or obtain an unfair advantage over other students.
* I will not be a party to any violation of the ISB Honour Code.
* I will personally uphold and abide, in theory and practice, the values, purpose and rules of the ISB Honour Code.
* I will report all violations of the ISB Honour Code by members of the ISB community.
* I will respect the rights and property of all in the ISB community.
* I will abide by all the rules and regulations that are prescribed by ISB.

**Note:** Lack of awareness of the ISB Honour Code is never an excuse for a violation. Please go through the Honour Code in the student handbook, understand it completely. Please also pay attention to the following points:

* Please do not share your assignment with your fellow students under any circumstances if the Honour Code scheme prohibits it. The HCC considers both parties to be guilty of an Honour Code violation in such circumstances.
* If the assignment allows you to refer to external sources, please make sure that you cite all your sources. Any material that is taken verbatim from an external source (website, news article etc.) must be in quotations. A much better practice is to paraphrase the source material (it still must be cited).

*(Please start writing your assignment below)*

1. *What are the problems Bigbasket is trying to address? Please be as specific as possible.*

A new issue had come into light for Big Basket which is being addressed by its tech team.

* **Too many items to remember – Forgetful for customer** – It came to light that big basket customers order grocery items in bulk quantity and many times forget to put a few items in their cart and buy them. This makes them order twice in a single day or go to other local shops to complete the forgotten items.
* **Problem of large no of items on small devices** – Big Basket has more than 1000 brand sand 30,000 plus items. It gets difficult for the customers to scroll through all the items on a weekly basis because of which they find it difficult to select all desired items and want an automated process.

1. *What is/are some of the fundamental differences in the recommender systems requirements between Bigbasket and other eCommerce companies such as Amazon or Flipkart?*

Ecommerce giants like Amazon or Flipkart have different recommender systems than Bigbasket as their business models and company objectives are different than Bigbasket’s.

Big Basket focuses on selling grocery items online whereas Amazon and Flipkart sell electronics, toys, apparel etc online.

Customers make purchases of daily use and similar kitchen and food items online from Bigbasket on a highly frequent basis, whereas on purchase frequency on Amazon and Flipkart are not that repetitive.

The recommender system of eCommerce websites will not work the best for Bigbasket platform.

1. *Discuss what sort of recommendation technique(s) is/are more appropriate for Bigbasket, in general, and why? Elaborate by connecting different methods and techniques with specific use cases.*

Given the differences in question 2, Bigbasket requires a different recommendation system from the ones which eCommerce websites use.

A recommender system which predicts the customer purchase behaviour and addresses their problem areas when making a purchase on the website.

Model for Issue 1 – Did You Forget

We can create an associative rule basis the transactional data of the customers. Based on the general data created on the platform, we can generate new choices and options for a new customer purchasing for the first time. And for the old customers, we can create a customized recommendation system that uses previous orders data to track and recommend what that particular consumer might like to buy in his next purchase.

Model for Issue 2 – Smart Basket

We can use Smart Basket model approach through Collaborative filtering to address this issue. In this model we can predict the purchase of a user based on the similar groups of previous purchases made by other customers. A smart basket model will ideally help in the predictive shopping and can capture the purchases records of a particular customer, and can recommend similar purchases to other customer who is following similar purchase pattern. This can reduce time of the customers of long time scrolling and searching of items and adding them to their carts.

1. *What sort of data challenges do you anticipate while building a recommendation engine for Bigbasket? Explore the data set and try applying different techniques on the dataset to identify the challenges that one would face while building the engine for Bigbasket.*

Challenges anticipated –

* The dataset consists of wrong items in the date column, that can create a hindrance of computing results based on time.
* Association rule will be used in computing recommendations for the model. With these challenges can arise when data is computed for every customer, making it difficult to track and sometimes can challenge on the accuracy of the recommendations made.
* In this only those items will be presented as a choice of purchase which will be bought by the other users. Any new category related items, for eg. kitchen cooking utensils which are not so frequently bought items, may not be recommended based on previous purchases.
* On Smart Basket model, new users can face difficulties as there wont be any previous transactional data to record and use to recommend items to purchase.

1. *What are the implementation and deployment challenges of a recommendation engine for Bigbasket?*

We may face a few challenges once the model is deployed –

* Since the customer base is quite large and transactional data per customer will even be larger in size, the system can face difficulty in smooth execution and can slow down.
* This can take a lot of time to load the data as well for the customers at user end.
* The model will work if the customer wants to order re ordered items or make a repeat purchase. If the customer wants to buy a completely new item not purchased before, this recommendation system will not work.

1. *Write an R script to generate five consumer-agnostic “good-quality” association rules sorted by their lift ratios? Please include the output (copy-paste/screenshot) ─ those five rules along with metrics that quantify the “goodness” of those rules ─ in your submission. What is the support of the first rule? Explain how it has been calculated for this rule. What is the confidence of the first rule? Explain how it has been calculated for this rule. What is the lift ratio of the first rule? Explain how it has been calculated for the rule. Based on the five rules you identified, suggest a couple of action plans that can benefit Bigbasket.*

To approach this question, I have created consumer agnostic good quality association rules.

To read the data better, we combined the ‘Description’ and ‘Order’ columns into ‘Transactions’ data and then analysed it through apriori algorithm to generate association rule output.

The model provides recommendation to general shoppers with previous purchase data.

Support of the first rule is 13.1%

How it is calculated - %( or number) or transactions in which antecedent (IF) and consequent (THEN) appear in the data.

Confidence of first rule is 67.4%

How it is calculated - % of antecedent (IF) transactions that also have the consequent (THEN) item set, same as –

P(Consequent | Antecedent)

= P(C & A)/ P(A)

= #transactions with both antecedent and consequent / #transaction with antecedent item sets

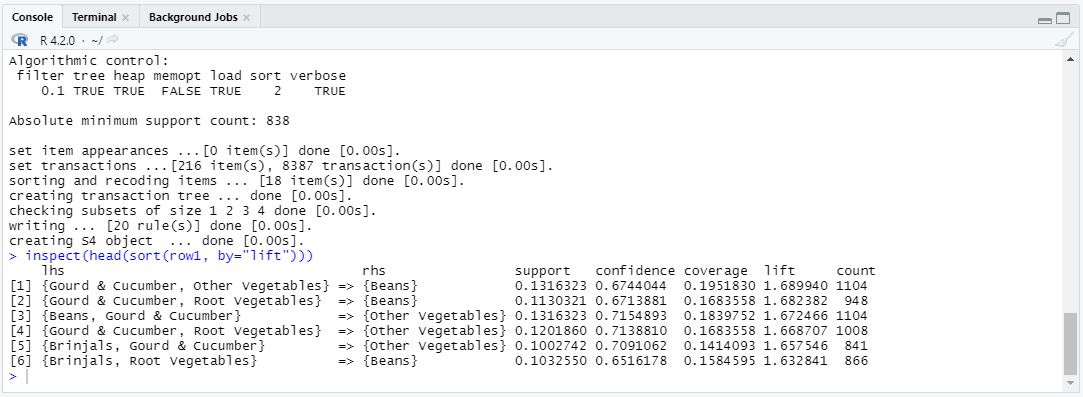
Lift ratio of first rule is as high as 1.69

How it is calculated – confidence/(benchmark confidence)

Benchmark assumes independence between antecedent and consequent –

P(antecedent & consequent) = P(antecedent) X P(consequent)

The output based on Association rule and with goodness metrics are shown below –



1. *Bigbasket is interested in introducing a “Did you forget?” feature to identify items a customer may have forgotten. Discuss how this feature can be created following data-driven approaches.*

“Did you forget” model will be used and beneficial for customers who make large order purchases and sometimes forget to add a couple of items in their order purchase.

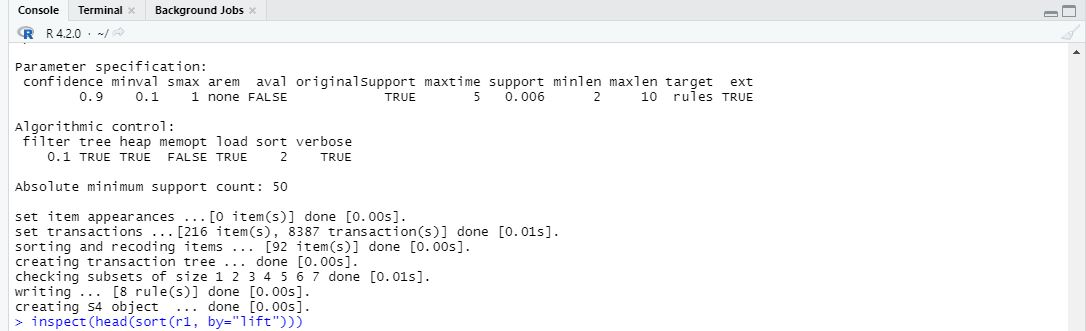
2 types of association rule can be created – one based on all transactions and customer-based transaction.

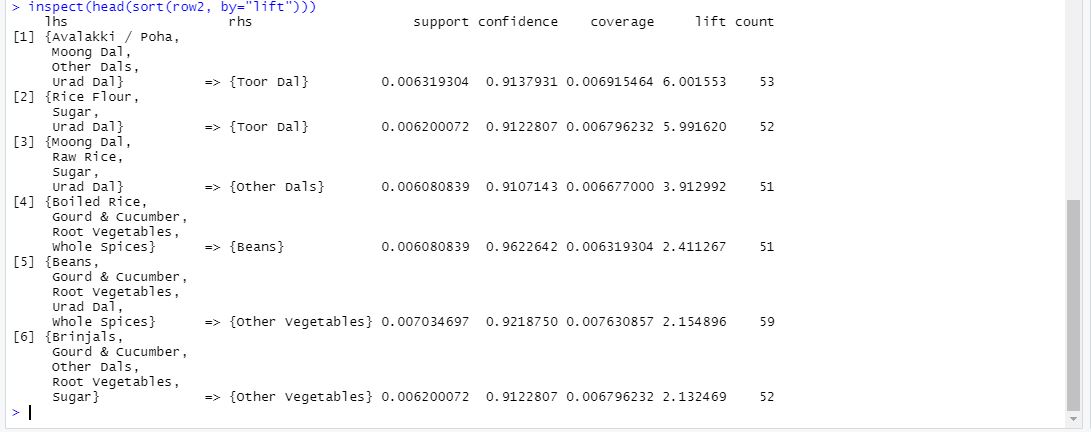
Both the association rules model will work basis on the transactions recorded in the data and will recommend the users if they forgot any items to be added in their purchase. Customer based transaction will be more personalized than all transaction-based model as it will recommend a particular user regarding his specific items to be added that he might have purchased in the past.

1. *Write an R script to do the following: Pick a customer of your choice. Now, present an example of product sets A and B such that when a shopper has added product set A in his/her cart, the recommendation engine should suggest product sets B as “Did you forget items?” Support your answers using output generated by your scripts.*

For the analysis, we combined the ‘Description’ and ‘Order’ columns to create a combined ‘Transactions’ column and created Association rule on the basis for it.

Here is the code output -





1. *How do we find similarities between users based on what customers buy? Can collaborative filtering be used to create values given this dataset? Present a use case where finding similarities between user baskets can be helpful to Bigbasket.*

Yes, Collaborative filtering can be used for this data set to generate user recommendations based on what other similar users are purchasing in the similar categories.

Use case – for eg a customer A has bought ingredients to make a cheese pizza and has bought items like pizza sauce, cheese, onion, capsicum & pizza base. This data has been recorded in the system.

When another consumer B will go on the website and select pizza sauce and pizza base to add in his cart, the recommendation system will predict his other ingredient choices and will recommend him items purchased by customer A like cheese, capsicum and onion.

This system will help the customers to not scroll through all the items in their category lost and can also help them remember all the items they might need for their particular purchase.

1. *How do you propose to build a Smart Basket for a customer on Bigbasket?*

As mentioned above, the Smart Basket model can give recommendations based on users’ past purchases and wat they would want to buy.

This can be achieved either by association rule or collaborative filtering. Better would be collaborative filtering in which the recommendation system can be built using past purchase behaviour of the customers. In this, similar recommendations will be provided base on what other similar category purchase have been made by other customers. User based collaborative filtering would be beneficial for this.