Assignment - Association Rule Mining

Pages

Problem Satatement

Note: The deadline for the submission of this individual assignment is 3 Jan, 2017, 11:59 PM. The 2nd deadline with 30% penalty is 9 Jan 2017.

Business Problem

You are already familiar with the "Global Mart" online store. It's an online store that caters to customers from across the globe. As the marketing manager of the store, you want to figure out the most frequently occurring combination of the items that are bought together. This would enable you to recommend the related items to a customer, once he makes a purchase in the store.

Data Understanding

You can download the Global Mart dataset and its data dictionary from the link below.

Dataset
*file_download*Download

Data Dictionary
*file_download*Download

As you would see, each row of the dataset represents an item of the order. However, the *Order ID* is not unique. Thus, the different items ordered at a time figure in different rows with the same Order ID. The number of unique products is too large a number to provide meaningful insight. Thus, the most relevant attribute to analyse would be the "Sub-Category" of the products.

Data Preparation

You have already seen that market basket analysis can be done on the transaction level data, where each row represents the items that are bought in a single transaction. To be able to use the "arules" package for association rule mining, you need to convert your data into transactions format. For that, you can either use as() function or read.transactions() function. It is advisable that you explore both these functions and their requirements and utility before proceeding further.

Data Mining & Evaluation

Once, you have the transaction level data, where each row represents the items bought in a single transaction, you can explore the association rules from the data using apriori principle. Here you would want to find the most business-relevant association rules, while also maximising the support, confidence or lift. You can also put a cap on your transactions by considering only those which have more than a threshold itemsets.

Checkpoints

Checkpoint 1: (Data Understanding & Data Preparation) (30% marks)

- Transform the original dataset into transaction level dataset (10%)
- Convert the transaction level data into transaction format using "arules" package (20%)

Checkpoint 2: (Association Rule Mining) (30%)

- Mine association rules from the data (15%)
- Optimise the minimum support, confidence, lift threshold level or the minimum floor on the number of items required in a transaction to qualify for consideration (15%)

Checkpoint 3: (Rule Relevance/Evaluation) (40%)

- The numerical value of the support, confidence and the lift level for the itemsets/rule (10%)
- How logical/relevant/viable are the rules from the store's point of view (10%)
- Explain and analyse the business implications of the rule (20%)

Submission

Important Note: All your code has to be submitted in one main .R file. Please make sure to rename your R script with "**Roll_Number_main.R**".

Also, you have to prepare a short PPT to present the results of your model. This should briefly describe the important results and recommendations. Name the PPT "Roll_Number_main.pptx".

An ideal solution must satisfy following criteria:

	Evaluation Parameters	
S.No	**Checkpoint**	**Evaluation Parameter**
1	Data understanding & data preparation	Efficiency of the code, transforming the order level data to transaction level data and reading into transaction format
2	Data mining	Efficiency of the code and adherence to the logical flow of the steps in association rule mining
3	Model evaluation / Interpretation	The measure of the support, confidence and the lift of the relevant itemsets/rules and coming up with most logical, relevant and coherent interpretation and implication of the result in context of the business problem

In total, you have to upload the two files as **one zip file named "Roll_no.zip".**The zip file should contain the **main R file, one PPT.**

Association Rule Mining - Assignment*today*3rd January 2017

Association Rule Mining - Assignment

The deadline for submission of this case study is **3 Jan 2017, 11:59 PM**. For submissions obtained within 1 week of the deadline, there will be a 30% penalty. Submissions beyond 1 week of the deadline will not be accepted.