**Association Rules**

**Instructions:**

Please share your answers filled in-line in the word document. Submit code separately wherever applicable.

Please ensure you update all the details:

**Name: \_\_\_\_\_\_\_\_\_\_\_\_\_ Batch ID:** \_\_\_\_\_\_\_\_\_\_\_

**Topic: Association Rules**

**Guidelines:**

**1. An assignment submission is considered complete only when the correct and executable code(s) is submitted along with the documentation explaining the method and results. Failing to submit either of those will be considered an invalid submission and will not be considered a correct submission.**

**2. Ensure that you submit your assignments correctly and in full. Resubmission is not allowed.**

**3. Post the submission you can evaluate your work by referring to the keys provided. (will be available only post the submission).**

Hints:

* Business Problem
* **What is the business objective?**
* **Are there any constraints?**
* Work on each feature of the dataset to create a data dictionary as displayed in the below image**:**



* Data Pre-processing
* Data Cleaning, Feature Engineering, etc.
* Model Building

4.1 Application of Apriori Algorithm

* Build the most frequent item sets and plot the rules
* Work on Codes

5. Deployment

5.1 Deploy solutions

6. Write about the benefits/impact of the solution - in what way does the business (client) benefit from the solution provided?

**Problem Statement: -**

Q1. Kitabi Duniya, a famous bookstore in India, was established before Independence, the growth of the company was incremental year by year, but due to the online selling of books and widespread Internet access, its annual growth started to collapse. As a Data Scientist, you must help this heritage bookstore gain its popularity back and increase the footfall of customers and provide ways to improve the business exponentially to an expected value at a 25% improvement of the current rate. Apply the pattern mining techniques (Association Rules Algorithm) to identify ways to improve sales. Explain the rules (patterns) identified, and visually represent the rules in graphs for a clear understanding of the solution.

**1.) Data: Books.csv**



**Business Objective:**

**Business Constraints:**

**Success Criteria:**

**Business Success Criteria:**

**ML Success Criteria:**

**Economic Success Criteria:**

**Questions to Ignite your Thinking process:**

Q1. Which library/package is used for the Association rules algorithm?

The mlxtend library in Python is commonly used for implementing the Apriori algorithm, which is a popular algorithm for mining association rules. The mlxtend library provides implementations for various machine learning algorithms, including association rule mining.

Q2. Which functions are used in the Association rules algorithm?

1.find\_frequent\_itemsets: This function is used to find frequent itemsets in a dataset. It takes as input the dataset and a minimum support threshold and returns the frequent itemsets.

generate\_rules: This function is used to generate association rules from the frequent itemsets. It takes as input the frequent itemsets and a minimum confidence threshold and returns the association rules.

Q3. What is the keyword used to import any package to the Python session’s memory?

import numpy as np

Q4. What type of data is usually worked in Association rules?

Association rule mining is typically applied to transaction data, where each transaction consists of a set of items. This type of data is often represented in a tabular format, with transactions as rows and items as columns.

Q5. Association rules are also named as

Association rules are also known as "market basket analysis" or "affinity analysis" in the context of retail and transaction data analysis.

Q6. What is the IF part called in an Association rule?

In an association rule, the "IF" part is called the antecedent. It represents the condition or itemset that, when present in a transaction, implies the presence of another item or items in the same transaction (the "THEN" part, or consequent).

Q7. What is the THEN part called in an Association rule?

In an association rule, the "THEN" part is called the consequent. It represents the item or items that are likely to be found in a transaction if the antecedent (the "IF" part) is present in that transaction.

Q8. In which sector is the Association rules algorithm mainly used?

The Association rules algorithm is mainly used in the retail sector, especially in market basket analysis.

Q9. What do slotting fees mean?

Slotting fees, also known as slotting allowances, are fees charged by retailers to manufacturers or suppliers for the placement of their products on store shelves or in prominent store locations. These fees are typically paid as a condition for allowing the supplier's product to be stocked and sold in the retailer's stores.

Q10. How is Support calculated in the Association rules algorithm?

Support is a measure used in the Association rules algorithm to indicate the frequency or popularity of an itemset in a dataset. It is calculated as the proportion of transactions in the dataset that contain the itemset.

Mathematically, the support of an itemset X is calculated as:

Support(X)=Number of transactions containing /Total number of transactions

​

Q11. What is the drawback of Support in Association rules algorithm?

One drawback of using support as a measure in the Association rules algorithm is that it does not take into account the number of transactions that contain the antecedent (IF part) of a rule. This can lead to misleading results, especially when dealing with rare items or itemsets.

Q12. To remove the infrequent items from data which algorithm is used?

To remove infrequent items from data, the Apriori algorithm is commonly used. The Apriori algorithm is an association rule mining algorithm that works by iteratively finding frequent itemsets in a dataset and then using those itemsets to generate association rules. Items that do not meet a specified minimum support threshold are considered infrequent and are pruned from the dataset.

Q13. The drawback of Support is captured by

The drawback of Support is captured by the fact that it does not consider the number of transactions that contain the antecedent (IF part) of a rule. This can lead to misleading results, especially when dealing with rare items or itemsets.

Q14. How is confidence calculated in the Association rules algorithm?

Confidence is a measure used in the Association rules algorithm to indicate the likelihood that a rule is true. It is calculated as the proportion of transactions that contain the antecedent (IF part) of the rule and also contain the consequent (THEN part) of the rule.

Mathematically, the confidence of a rule A -> B is calculated as:

Confidence(A→B)=Support(A∪B)/Support(A)

​

Q15. What is the drawback of Confidence in Association rules algorithm?

The drawback of Confidence in the Association rules algorithm is that it can be misleading when the support for the antecedent (IF part) is low.

Q16. How is the Lift ratio calculated?

The Lift ratio is a measure used in the Association rules algorithm to indicate how much more likely the consequent (THEN part) of a rule is to be present in transactions containing the antecedent (IF part) compared to its presence in all transactions.

Mathematically, the Lift of a rule A -> B is calculated as:

Lift(A→B)= Support(A∪B)/Support(A)×Support(B)

​

Q17. What is the threshold value of the Lift ratio of a rule to declare it as a good Association rule?

**There is no fixed threshold value for the Lift ratio to declare a rule as a good association rule, as it depends on the specific dataset and the context of the analysis. However, a Lift value greater than 1 is generally considered indicative of a meaningful association between the antecedent and consequent of a rule. A higher Lift value indicates a stronger association.**