Association Rules

The Objective of this assignment is to introduce students to rule mining techniques, particularly focusing on market basket analysis and provide hands on experience.

**Dataset:**

Use the Online retail dataset to apply the association rules.

**Data Preprocessing:**

Pre-process the dataset to ensure it is suitable for Association rules, this may include handling missing values, removing duplicates, and converting the data to appropriate format.

**Association Rule Mining:**

* Implement an Apriori algorithm using tool like python with libraries such as Pandas and Mlxtend etc.
* Apply association rule mining techniques to the pre-processed dataset to discover interesting relationships between products purchased together.
* Set appropriate threshold for support, confidence and lift to extract meaning full rules.

**Analysis and Interpretation:**

* Analyse the generated rules to identify interesting patterns and relationships between the products.
* Interpret the results and provide insights into customer purchasing behaviour based on the discovered rules.

# **Interview Questions:**

1. What is lift and why is it important in Association rules?
2. What is support and Confidence. How do you calculate them?
3. What are some limitations or challenges of Association rules mining?

Answers: <https://colab.research.google.com/drive/13drFsJg_hrwjUrfqW318GzZKFWEve3A4?usp=sharing>

1. Lift is a measure used in association rule mining to evaluate the strength and importance of a rule. It is defined as the ratio of the observed support of an item-set to the expected support if the items were independent.

Importance of Lift:

Indicator of Independence: A lift value of 1 indicates that X and Y are independent. A lift greater than 1 indicates a positive correlation, meaning that the occurrence of 𝑋 increases the likelihood of Y.A lift less than 1 indicates a negative correlation.

1. Support and Confidence are key metrics in association rule mining. Support measures the frequency of an item-set occurrence in the dataset, helping identify common patterns. Confidence measures the reliability of a rule, determining how often items in a rule appear in transactions containing the rule. Both metrics are calculated using the formula:

Support(𝑋) = 𝑋 ∪ 𝑌  
, where 𝑋 is the number of transactions containing the item, and X is the total number of transactions.

1. Association rule mining is a powerful technique for discovering interesting patterns in data, but it has several limitations and challenges. These include computational complexity, scalability, redundancy, interpretability, support/confidence trade-off, handling rare item-sets, context sensitivity, threshold setting, data sparsity, handling continuous data &dynamic data.

Despite these challenges, association rule mining remains a valuable technique for discovering interesting patterns in data, and advances in algorithms and techniques continue to address these limitations. Effective pre-processing, parameter tuning, and post-processing of the rules are essential for its effectiveness and application.