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?

Answer: Lift is a measure of how much more often the antecedent and consequent of a rule occur together than we would expect if they were statistically independent. It is calculated as the ratio of the observed support to the expected support under independence. Lift value greater than 1 indicates that the antecedent and consequent are more likely to occur together, suggesting a strong association. Lift helps in identifying meaningful relationships between items in a dataset, which is crucial in market basket analysis and recommendation systems.

1. What is support and Confidence. How do you calculate them?

Answer:

* Support is the proportion of transactions in a dataset that contain the items in a rule. It indicates the popularity or frequency of an itemset in the dataset. It is calculated as the number of transactions containing the items in the rule divided by the total number of transactions.
* Confidence is the conditional probability that a transaction containing the antecedent will also contain the consequent. It indicates the strength of the association between the antecedent and consequent. It is calculated as the number of transactions containing both the antecedent and consequent divided by the number of transactions containing the antecedent.

1. What are some limitations or challenges of Association rules mining?

Answer:

* Sparse Data: Association rule mining can be challenging when dealing with sparse datasets where most itemsets have low support. This can lead to a large number of rules with low confidence, making it difficult to identify meaningful patterns.
* Multiple Comparison Problem: When mining association rules from large datasets, there can be a huge number of possible rules. This increases the chance of finding spurious associations due to multiple comparisons.
* High Dimensionality: As the number of items in a dataset increases, the number of possible itemsets and rules also increases exponentially. This can make it computationally expensive to find and analyze all possible rules.
* Interpretability: While association rules can reveal interesting patterns in data, interpreting these rules and translating them into actionable insights can be challenging, especially for complex rules involving multiple items.