NAME: KSHITIJ VINOD SALI

CLASS & DIV: TE-A

ROLL No.: 35059

ML

UA-05

Title: Association Rule Learning.

Problem Statement: Download Market Basket Optimization dataset from below link:

https://www.kaggle.com/hemanthkumar05/market-basket-optimization

This dataset comprises the list of transactions of a retail company over the period of one week. It contains a total of 7501 transaction records where each record consists of the list of items sold in one transaction. Using this record of transactions & items in each transaction, find the association rules between items.

There is no header in the dataset & the first row contains the first transaction, so mentioned header = None here while loading dataset.

Follow following steps:

a) Data preprocessing.

b) Generate the list of transactions from dataset.

c) Train Apriori algorithm on the dataset.

d) Visualize the list of rules.

e) Generated rules depend on the values of hyper parameters. By increasing the minimum confidence value & find the rules accordingly.

Objective: Association rule learning is a rule-based

machine learning method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using some measures of interesting-ness.

Theory:

1) Association Rule Learning:

→ It is a type of unsupervised learning technique that checks for the dependency of one data item on another data item & maps accordingly so that it can be more profitable. It tries to find some interesting relations or associations among the variables of the dataset. It is based on different rules to discover the interesting relations between variables in the database.

The association rule learning is one of very important concept of machine learning, & it is employed in market Basket analysis, web usage mining, continuous production, etc. Here market basket analysis is a technique used by various big retailers to discover the association between items. We can understand it by taking an example of a supermarket, as in a supermarket, all products that are purchased together are put together.

2) How does Association Rule Learning work?

→ Association rule learning works on the concept

of If and Else statement, such as if A then B.

$$\boxed{\text{If } A} \longrightarrow \boxed{\text{Then, B}}$$

Here, the "If" element is called antecedent, and then the statment is called as consequent. These types of relationships where we can find out some association or correlation between two items is known as single cardinality. It is all about creating rules, and if the number of items increases, then cardinality also increases accordingly. So, to measure the associations between thousands of data items, there are several metrics. These matrics are:

• Support
• Confidence.
• Lift.

i) Support: It is the frequency of A or how frequently an item appears in the dataset. It is defined as the fraction of the transaction T that contains the itemset X. If there are x datasets, then for transactions T, it can be written as:

$$\text{Supp }(x) = \frac{\text{Freq }(x)}{T}$$

ii) Confidence: It indicates how often the rule has been found to be true, or how often the items X and Y occur together in the dataset

when the occurance of x is already given. It is the ratio of x is already transaction that contains X and Y to the number of records that contains X.

$$Confidence = \frac{Freq(X,Y)}{Freq(X)}$$

iii) Lift: It is the strength of any rule, which can be defined as below formula:

$$Lift = \frac{Supp(X,Y)}{[Supp(X)] \times [Supp(Y)]}$$

It is the ratio of observed support measure & expected support if x & Y are independent of each other. It has 3-possible values:

- If Lift = 1: The probability of occurance of antecedent & consequent is independent of each other.
- If Lift > 1: It determines the degree of which the two itemsets are dependent on each other.
- If Lift < 1: It tells us that one item is a substitute for other items, which means one item has a negative effect on another.

3) Types of Association Rule Learning.
→ Association rule learning can be divided into three algorithms:

i) Apriori

ii) Eclat

iii) F-P Growth Algorithm.

Conclusion: Association Rule Mining Collects Interest-ing Association and Correlation Relationships among large sets of data items. The association rule shows attribute value conditions that occur frequently together in a given dataset. A simple example of association rule mining is Market Basket Analysis.