**Assignment 2:**

**Title: Sequence Pattern mining**

**About Sequence Pattern Mining**

In the sequence pattern mining we are exploring patterns shared among transactions where a set of itemsets occurs sequentially.

**Activities Performed throughout the code has been listed below**

Here, i have used sequential pattern mining algorithm,

I have used **SPADE, to mine frequent sequential patterns.**

**About SPADE**

**Sequential Pattern Discovery using Equivalence classes (SPADE) algorithm, which uses the**

**characteristics of a vertical database to perform an intersection on an ID list with an**

**efficient lattice search and allows us to place constraints on mined sequences.**

1. First, i have loaded temporal Information transactions into the variable **sales\_data.tran.**

**Creation of sales\_data.tran**

**In the first method,**

**{**

**sales\_data.tran<-as(tran\_data,"transactions")**

**transactionInfo(sales\_data.tran)$sequenceID <- sales\_data$sequence**

**transactionInfo(sales\_data.tran)$eventID<-sales\_data$TIME**

**}**

**I have created a list of transactions, and assigned a transaction ID for each transaction.**

**I have used (as) function to transform the list data into a transaction dataset.**

**After that added eventID and sequenceID as temporal information.**

**sequenceID is the sequence that the event belongs to, and eventID indicates when the**

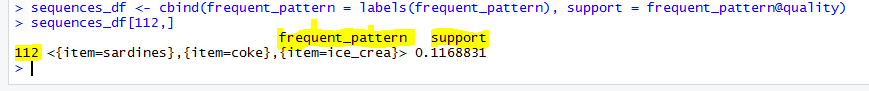
**event occurred.**

**So, after generating transactions with temporal information, i have used this dataset for frequent sequential pattern mining using cSpade.**

1. I used **cspade function** with support over **0.045** to generate frequent sequential patterns in sequences format.
2. After this, used **SUMMARY** in order to extract information such as the most frequent items, sequence size distribution, the summary of quality measures, and mining info.
3. At last I have transformed the generated sequence information back to data frame format, so that i can examine the sequence and support of frequent sequential patterns with support over 0.045.
4. Lastly I have found the access pattern of a customer from the generated patterns. A retail industry can inspect the patterns to know how to improve/keep their inventory as per the customer sequence of buying products.

**OUTPUT Results:**

**Output of the sequence pattern for** **Rule 112**



**For example, in the above output**

Here, rule 112 says that “**If a customer’s first purchase is "sardines", his second purchase would be "coke", followed by his third purchase of "ice\_cream”** which is **frequent for 11% of customers.**

It means that each of unique sequences happened on the same **Week.**