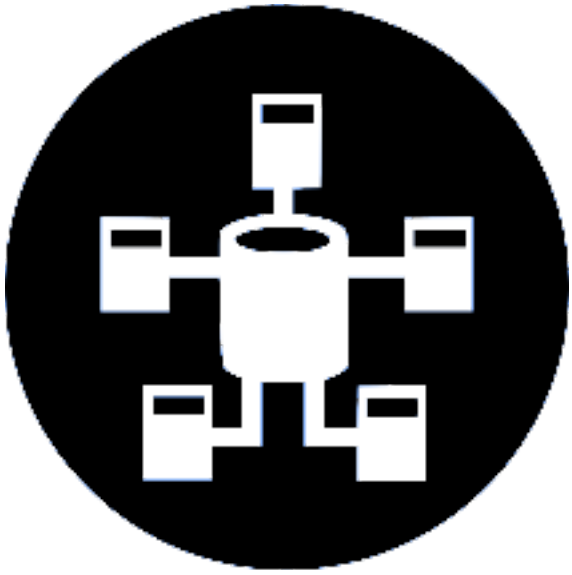


Association Rule Mining

(Assignment-3)

Name: Sarthak Singhal

Roll No: 20171091

Technology Used: Python

EXPERIMENTS TO BE PERFORMED

TASK-1

A. **Remove stop words and get for each line one transaction.(Be careful while removing the stop words)**
B. **Get Frequent Itemsets and association rules with the transaction keeping low minsupport value**

TASK-2

A. **Now remove from transactions "Knock, knock" and "who is there?" That is more common words occurring in 90% or more transactions**
B. **Now get Frequent Itemsets and association rules from these transactions.**

TASK-3

A. **Come up with few knock-knock jokes , add them to your data so that one in-frequent item set becomes frequent.(The jokes doesn't have to be funny! or semantically correct)**

PARAMETERS:

Min support=0.1;Min confidence=0.1;Min lift=1;Min length=1;Max length=2

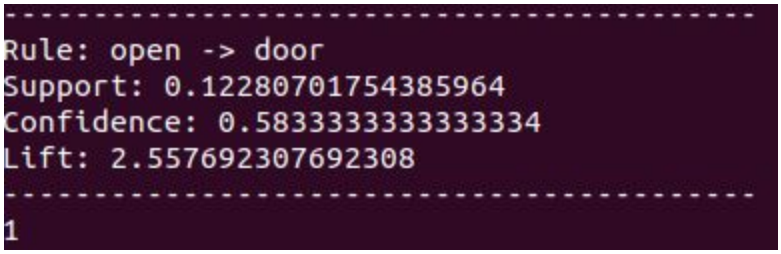
ANS-1

```
----------------------------------------
Rule: knock -> adore
Support: 0.10526315789473684
Confidence: 1.0
Lift: 1.0
----------------------------------------
----------------------------------------
Rule: door -> knock
Support: 0.21052631578947367
Confidence: 1.0
Lift: 1.0
----------------------------------------
----------------------------------------
Rule: open -> door
Support: 0.12280701754385964
Confidence: 0.5833333333333334
Lift: 2.557692307692308
----------------------------------------
----------------------------------------
Rule: open -> knock
Support: 0.22807017543859648
Confidence: 0.22807017543859648
Lift: 1.0
----------------------------------------
----------------------------------------
Rule: open -> door & knock
Support: 0.12280701754385964
Confidence: 0.5833333333333334
Lift: 2.557692307692308
----------------------------------------
5
```

CONCLUSIONS

- If we look at count of items present in all association rules, we will find that some of the items have very high count and others have very low count.
- Items which occur in more than 90% of transactions are involved in most of the association rules.
- Items which are present in almost all the transactions are involved in association rules which does not give us a true picture of the dataset that how the items are linked to each other.

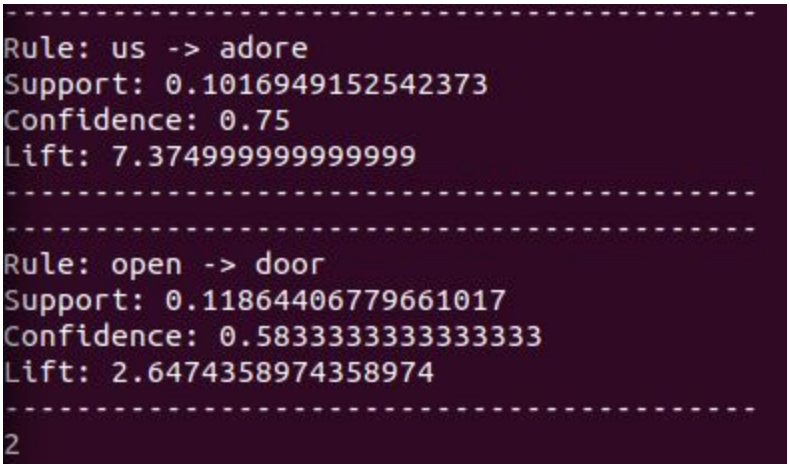
ANS-2



CONCLUSIONS

- Since most of the association rules consisted of items which occurred in 90% of the transactions, we get very less rules when all those items were removed.
- Since data is random, support of items is very less.
- The items which we get as frequent give us a true picture of relations in the data.

ANS-3



New transactions added:

- Knock knock. Who's there? Adore. Adore who? Adore is flying above us.
- Knock knock. Who's there? Adore. Adore who? Adore is gonna hit us so we must climb up.