**Birla Institute of Technology and Science-Pilani, Hyderabad Campus**

# SEMESTER-II 2018-2019


## Data Mining (CS F415)

## Association Rule Mining

BY

Garvit Jain 2016A7PS0080H
Aman Mehta 2016A7PS0066H
Ayushi Srivastava 2016A1PS0587H

UNDER THE SUPERVISION OF

Mrs. Aruna Malapati

# Dataset

**Groceries Market Basket Dataset**

- Number of transactions: 9835
- Number of unique items: 169

# Pre-processing done on data

Groceries.csv file was read transaction by transaction and each transaction was saved as a list. A mapping was created from the unique items in the dataset to integers so that each item corresponded to a unique integer. The entire data was mapped to integers to reduce the storage and computational requirement. A reverse mapping was created from the integers to the items, so that the item names could be written in the final output file.

# Formulas Used

- Confidence (X -> Y) = support (X U Y) / support (X)
- Support (X, Y) = support-count (X, Y) / total dataset size

## Results for different for values of support and confidence

| Confidence/Support | No. of Frequent Itemsets | No. of rules |
|---|---|---|
| High confidence (MIN_CONF=0.5) High support count(MINSUP=60) | 725 | 65 |
| Low confidence (MIN_CONF=0.1) High support count(MINSUP=60) | 725 | 1219 |
| High confidence (MIN_CONF=0.5) Low support count(MINSUP=10) | 11390 | 4484 |
| Low confidence (MIN_CONF=0.1) Low support count(MINSUP=10) | 11390 | 34497 |

# Redundant rules:

We kept a support count = 10 and confidence = 0.1 to get 12 non-closed frequent itemsets from which we found 49 redundant rules. Some of which are:

[These are in format X(X_sup)->Y(Y_sup), confidence]

1. rice(75)->sugar(333) , 0.16
2. brown bread , whipped/sour cream(46)-> pip fruit, 744, 0.24
3. pip fruit, whipped/sour cream(91)->brown bread(638), 0.12

# Observation:

Apriori takes significantly more time than Fp Tree algorithm as FP tree algorithm uses divide and conquer approach.

**Time taken for frequent item generation(Apriori) for Minimum Support=10**: 25.04 sec

**Time taken for frequent item generation(FP Growth) for Minimum Support=10**: 3.74 sec

Most of the rules we generated have a common item (whole milk and other vegetables) on the consequent side. This happens when any item is very frequent in the transactions. This can be avoided by using lift instead of confidence.

**Lift (X -> Y) = support (X U Y) / support (X) * support (Y)**