**Assignment 01: Apriori and Fp Tree Algorithm**

**Team Members**: Debanjan Ghatak(**2017MCS2079**),Saurav Kumar(**2017MCS2103**),Rakesh Raushan(**2017MCS2092**)

**1)** **Files that we have submitted:**

a) **Apriori Folder** which contains the code for **Apriori Algorithm**.

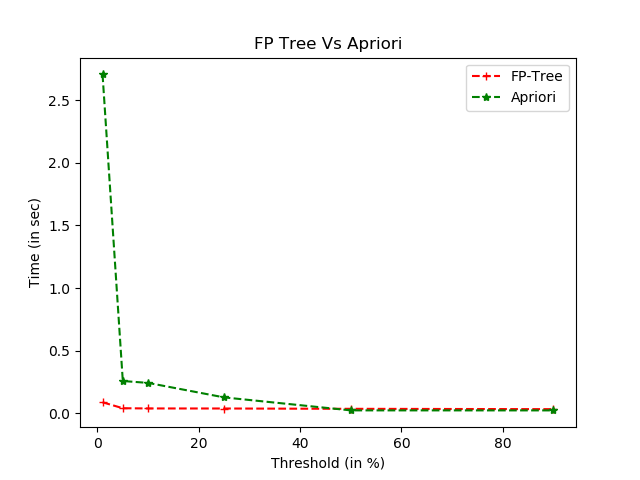
b) **FP\_Tree Folde**r which contains the code for **Fp Tree**.

c) **FP\_Tree Folder** which contains code for plotting graph for **Fp Tree**.

d) A **compile.sh** file which compiles our code.

e) A **2017MCS2092.sh** file which executes commands for question a,b and c.

f) **plot.py** file which generates the plots for **Apriori** and **FP Tree** Algorithm.

**2) Part B Plot**

**2) Part B Explanation**.

We observed that for **lower** support thresholds(**1%,5%,10%,25%**),**FP Tree** algorithm works much **faster** compared to **Apriori** Algorithm.The reason is that for lower support thresholds, the number of **frequent itemsets** are quite **many**, so the Apriori Algorithm goes on scanning the transaction database **multiple times** after creation of **candidate itemsets** and on the other hand the **FP Tree** Algorithm makes only **two** database scans.Since we know **database scans** consume **maximum time** so the FP Tree algorithm works **significantly faster** for **lower** support thresholds.

As we kept on **increasing** the support thresholds we observed that FP Tree and Apriori Algorithm almost had **similar** timings.For support thresholds(**50% and 90%**),we observed that **Apriori** works **faster** compared to **FP Tree**. The reason is that the number of frequent itemsets for **high support thresholds** are very **few**, so it happened that there were **no two sized candidate itemsets** generated and **Apriori** Algorithm **terminated** with only **one database scan**.On the other hand **FP Tree** made **two database scans** as there were some **one sized frequent itemsets**.

**In general , FP Tree Algorithm works significantly faster for low support threshold values whereas Apriori Algorithm tends to work faster for very high threshold values.**

**3) How to execute our files:**

**I)** Run the **compile.sh** file and then run the **2017MCS2092.sh** file for executing Apriori and FP Tree Algo at different support thresholds.

**II)** Run only **2017MCS2092.sh** file with **–plot** command to generate the plots for FP Tree and Apriori Algo.Two files named **fp\_time.txt** and **a\_plot\_time.txt** are also created which contains the respective support threshold execution times.