COL761 Assignment 1

Harsh Vardhan Jain   Lovish Madaan   Nikhil Goyal

2015CS50102   2015CS50286   2015CS50287

August 23, 2018

1 Submitted Files

- **apriori.cpp**: Apriori Implementation

- **fptree.cpp**: FP Tree Implementation

- **plot.py**: Python script for plotting the graph of running time v/s support threshold

- **compile.sh**: Compile script for both implementations

- **2015CS50286.sh**: Bash script as specified in the statement

2 Explanation for Plots

- Apriori Algorithm repeatedly scans the transaction database to count the support of the candidates. Total number of scans of the database - O(k) where k is the largest size occuring in the frequent itemsets. FP Tree on the other hand only scans the transaction database twice so it runs faster.

- When the support threshold decreases, the length of the largest size frequent itemset increases that is why initially Apriori runs better than FP Tree(because no candidate set is generated in Apriori so it finishes its work in 1 scan only).

- Moreover, with increase in largest size in frequent itemsets, Apriori needs to generate more and more candidates thus increasing its time complexity.

- Apriori is kind of generation and test algorithm, it even generates the candidates not occuring in the database and then checks for their presence whereas FP Tree is fragment growth - it only expands the patterns existing in the database in a DFS fashion.