Assignment 1

Instructions

1) Do this assignment individually.
2) Each student will have to implement two standard data mining algorithms.
3) Discussion is encouraged, but copying code from your partners/internet/any other source is not appreaciated. Turnitin will be used to detect such cases and the guilty will be penalised.
4) Use any programming language for implementation – C, C++, Java, Python (or any other language of your choice)
5) Submit two files on Moodle, one each for each algorithm. Please mention the name of the algorithm as descriptions inside your code.
6) A brief description of the dataset to be used is given here: https://wiki.csc.calpoly.edu/datasets/wiki/ExtendedBakery
7) The datasets are available here: http://www.cse.iitd.ac.in/~prajna/COL868/Assignment1/dataset1.csv
8) http://www.cse.iitd.ac.in/~prajna/COL868/Assignment1/dataset2.csv
9) Evaluation will be held by taking demos. The input to your code will be a support value (k) and the output should be the list of item sets that have a support > = k.
10) The due date is 29 August, 2016, 11:55 PM, IST.

Statement

1) Implement the following data mining algorithms in the programming language of your choice:
    1. Apriori
    2. FPGrowth
    3. ECLAT
    4. Pincer Search
    5. Border
    6. DIC

2) The assignment of algorithms to different students is given below:

Student	Apriori	FPGrowth	ECLAT	Pincer Search	Border	DIC
NITESH SINGH	√	√				
AJITA SHREE			√	√		
KUNAL KISHOR					√	√
MILAN BHANJIBHAI KATHROTIA	√		√			
VAMSI YALAVARTHI		√		√		
AKSHAY SURESH BHAT			√		√	
SAHIL YADAV				√		√
SHUBHAM SAGAR	√			√		
SURYAKANT PANDEY		√			√	
VISWA TEJA GAJULAVARTHY			√			√
EDUBILLI AVINASH	√				√	
NIKHIL KUMAR		√				√
ABHISHEK KUMAR	√					√
AYUSH VERMA		√	√			
HIRULKAR ANKET PRAKASH				√	√	