

Data Mining
Programming Assignment 3
Deadline : 22nd October , 11:59 PM

Guidelines :

- Team size : 3 or 4
- Language Allowed : Java
- Attach screenshots of the output in the report
- Any instance of plagiarism will be dealt with a F grade for all team members .
- There is no partial marking for the Bonus Question (given on page 2).

Question : Implement the diverseFP algorithm from the paper titled “Discovering Diverse Frequent Patterns in Transactional Databases” .

Link for the paper : [Diverse Frequent Pattern Mining](#)
(Also included in the folder)

Link for the dataset to be used : [Mushroom DataSet](#)
(Also included in the folder)

The Concept hierarchy to be used is included in the file named “Hierarchy.txt” .

Format of the concept hierarchy :
Parent : [child 1, child 2 ,]

Child1 : [list of children]

Child 2 : [list of children]

And so on ...

The paper presents an algorithm to find frequent itemsets first and then apply the notion of diversity . For finding frequent patterns initially , use FPGrowth algorithm from SPMF library .

Link for SPMF library : [Java SPMF library](#)

Bonus Question :

Using the notion of diversity , set a threshold value and prune the tree while creating the FP tree . Thus , if a pattern does not satisfy a certain threshold value of diversity then there is no need to further extend the tree from that point .