DEPARTMENT OF COMPUTER SCIENCE
CPSC 473 - Introduction to Data Mining
Fall 2020

Assignment 2

Due Date: Nov 27, 2020 at 11:59PM

Notes:

- This assignment can be finished in Java, Python, C, or C++.
- Zip all your files (code files, readme file, test data files, and others if needed) together into one file named "CPSC473_a#_yourlastname_yourfirstname_student#.zip" (or .rar, or .7z, or .tar), then submit this package to blackboard.
- Since this assignment is an implementation of frequent pattern mining algorithm, which is the same as assignment 1, input and output requirements will also be the same as the assignment 1.

Assignment Description:

The main task of this assignment is straightforward:

- Implement an FP-growth algorithm

For this assignment, in the package that you submit to the blackboard, you should include your implementation in a directory named FPgrowth.

Detail requirements for implementation (mostly the same as assignment 1):

1. Your program should be able to read transaction database from a .txt file.
2. The format of transaction database file will be the same as assignment 1.
3. All possible input variables should be entered by command-line arguments.
    a. For example, if, 1) the program is named "fpgrowth", 2) the file path of database is "./testdata.txt", 3) the minimum support threshold is set to 50%. The command to run your program can be:
    ```
    java fpgrowth ./testdata.txt 50
    ```
4. You should provide a README file for program compilation and execution.
5. For the output, a console output simply prints out the total number of frequent patterns.
    a. E.g.:
    ```
    |FPs| = 9
    ```
    b. (Optional) You can also print out the execution time of your program.
6. The output requirements are the same as assignment 1.

DEPARTMENT OF COMPUTER SCIENCE
CPSC 473 - Introduction to Data Mining
Fall 2020

Mark Distribution and Bonus Marks:

- The FP-growth ➔ 100%
- Bonus Marks (Described below) ➔ 10%

About bonus marks:

- To get the bonus marks of this assignment, you need to document any special enhancement you included in the implementation. For example, you used some special implementation tricks to deal with candidate generation, which, can improve the performance compare with the traditional method.
- The document does not have to be long, as long as you clearly explain a) what you did, and b) why do you think this method can improve the performance.
- The format of the document does not have to be formal, an informal report format should be good enough.
- The more special enhancements you include in your implementation the better chance you will get full bonus marks.
- The document should be in PDF, and named as "BonusMarkDocumentation.pdf".
- Bonus marks will be added into your final grade calculation. So, yes, if you get it, your will have some extra marks to be added into your final grade.