# **Project 1**

**Name: Caleb Thian Jia Le 田家樂**

**Student ID: NN6114035**

Experiment on dataset 2022-DM-release-testdata-2.txt ,let

low min\_sup = 0.1

low min\_conf = 0.5

high min\_sup = 0.1

high min\_conf = 0.5

Record the time used by apriori and FP-tree under different combination of parameters mentioned, with the number of rules found respectively.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Time(s) | | Number of Rules Found | |
| min\_sup | min\_conf | Apriori | FP-tree | Apriori | FP-tree |
| Low | Low | 18.36 | 59.21 | 10287 | 1799 |
| Low | High | 18.31 | 59.35 | 3295 | 611 |
| High | Low | 0.14 | 58.90 | 2 | 2 |
| High | High | 0.14 | 59.11 | 2 | 2 |

Observation:

* Increasing min\_sup reduces the time used of Apriori, because more frequent patterns can be pruned in process, thus increase the speed for pattern generation.
* Time used of FP-tree does not be reduces because min\_sup and min\_conf does not reduces the candidate patterns generated, unlike Apriori
* But at the same time, number of rules found decreased in both algorithm as the threshold of support for rules is increased.
* Increasing min\_conf, with low min\_sup, will reduces the number of rules found

Experiment on dataset ibm-2021.txt, let

low min\_sup = 0.025

low min\_conf = 0.02

high min\_sup = 0.04

high min\_conf = 0.5

Record the time used by apriori and FP-tree under different combination of parameters mentioned, with the number of rules found respectively

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Time(s) | | Number of Rules Found | |
| min\_sup | min\_conf | Apriori | FP-tree | Apriori | FP-tree |
| Low | Low | 4.47 | 0.63 | 2086 | 1110 |
| Low | High | 4.45 | 0.63 | 2046 | 1103 |
| High | Low | 0.64 | 0.59 | 16 | 12 |
| High | High | 0.58 | 0.63 | 15 | 11 |

Observation:

* Similar as observations mentioned.
* The only different is the time-used for FP-tree is approximately less or equal to the time-used for Apriori, this should because to the difference of datasets.
* Thus the relative time used depends on dataset but not the parameters. We cannot conclude that Apriori is always slower or faster than FP-tree.

Experiment on Kaggle dataset basket\_analysis.txt, let

low min\_sup = 0.1

low min\_conf = 0.1

high min\_sup = 0.2

high min\_conf = 0.5

Record the time used by apriori and FP-tree under different combination of parameters mentioned, with the number of rules found respectively

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Time(s) | | Number of Rules Found | |
| min\_sup | min\_conf | Apriori | FP-tree | Apriori | FP-tree |
| Low | Low | 0.74 | 1.79 | 438 | 252 |
| Low | High | 0.75 | 1.81 | 97 | 68 |
| High | Low | 0.13 | 1.79 | 12 | 54 |
| High | High | 0.13 | 1.77 | 3 | 47 |

Observation:

* Similar as above observation
* The only things different from others is that FP-tree found more rules than Apriori when high min\_sup. It maybe because of the pruning strategies is different between these two algorithms, making the counting is slightly difference.