EECS 4412 Data Mining
Assignment 1 Report

Name: Chun Ho Li
Student ID: 214908800
Date: 22 Feb 2021

Question 1)
Part A

levelUp.py works by applying Apriori algorithm to mine frequent itemset for a given threshold. Before applying the pre-candidate optimization, the algorithm will join on the prefixes which is created randomly to create pre-candidates, then applying the Apriori property to finalize the candidates. This can lead to a huge set of pre-candidates' generation, result in a slower running time. Here is the result before optimization by testing with level 5 of the mushroom dataset.

```
#pre-candidates: 236859
#candidates:     172300
Lapsed time:     187.406
```

To achieve a more efficient algorithm, we sort the item in the itemset from least to most frequent, such that the pre-candidates' generation can be more optimized and less pre-candidates will be created. Hence, increasing the performance of the algorithm. Here is the result with optimization, the number of pre-candidates are greatly reduced, and the running time is slightly improved as a result.

```
#pre-candidates: 173877
#candidates:     172300
Lapsed time:     170.566
```

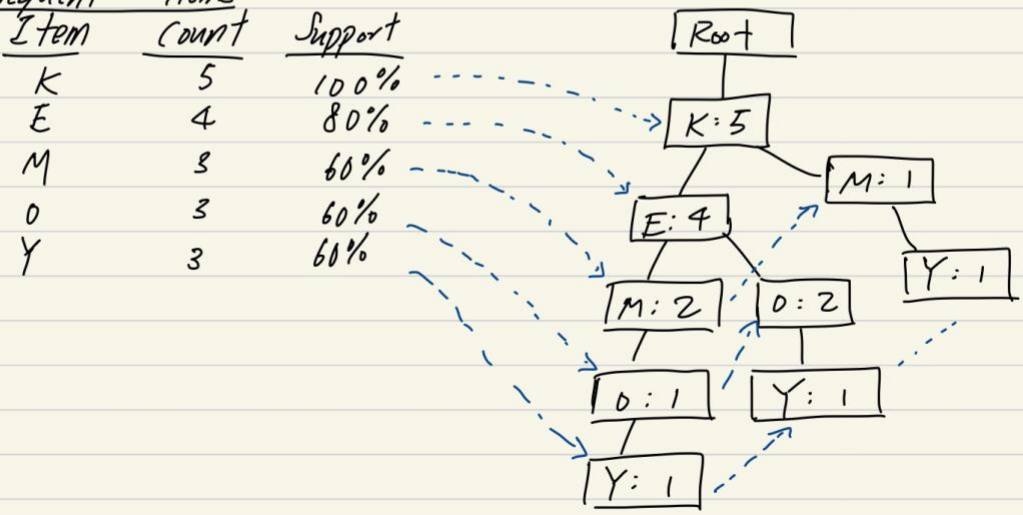
Question 2)
Part A)

A) Tranformed transactions itemsets

T1	{MONKEY}	→	{K E M O Y}
T2	{DONKEY}	→	{K E O Y}
T3	{MAKE}	→	{K E M}
T4	{MUCKY}	→	{K M Y}
T6	{COKIE}	→	{K E O}

Frequent items

Item	Count	Support
K	5	100%
E	4	80%
M	3	60%
O	3	60%
Y	3	60%



Part B)

B) Conditional Pattern-Bases and conditional FP-Tree

Items	conditional-pattern Base	Conditional FP-Tree
Y	KEMO:1, KEO:1, KM:1	{(K:3)} | Y
O	KEM:1, KE:2	{(KE:3)} | O
M	KE:2, K:1	{(K:3)} | M
E	K:4	{(K:4)} | E

Frequent Pattern Generated

{K, Y : 3}
{K, O : 3}, {E, O : 3}, {O, E, K : 3}
{K, M : 3}
{K, E : 3}

Association Rules Generation

{K} ⟶ {Y} [Support = 60%, confidence = 60%]
{Y} → {K} [Support = 60%, confidence = 100%]
{K} → {O} [Support = 60%, confidence = 60%]
{O} → {K} [Support = 60%, confidence = 100%]
{E} → {O} [Support = 60%, confidence = 75%]
{O} → {E} [Support = 60%, confidence = 100%]
{O, E} → {K} [Support = 60%, confidence = 100%]
{O, K} → {E} [Support = 60%, confidence = 100%]
{E, K} → {O} [Support = 60%, confidence = 75%]
{O} → {E, K} [Support = 60%, confidence = 100%]
{E} → {O, K} [Support = 60%, confidence = 75%]
{K} → {O, E} [Support = 60%, confidence = 60%]
{K} → {M} [Support = 60%, confidence = 60%]
{M} → {K} [Support = 60%, confidence = 100%]
{K} → {E} [Support = 80%, confidence = 80%]
{E} → {K} [Support = 80%, confidence = 100%]

After pruning for redundancy

$\{Y\} \longrightarrow \{K\}$
$\{E\} \rightarrow \{K\}$
$\{K\} \rightarrow \{E\}$
$\{O\} \rightarrow \{E, K\}$
$\{M\} \rightarrow \{K\}$