2020 ECT 作業二

對 adult.csv 資料集進行 Association Rule，並使用 Apriori 演算法，設定 confidence = 0.9、minimum support = 0.2，並回答以下問題：

No.	1: age Nominal	2: workclass Nominal	3: education Nominal	4: marital-status Nominal	5: occupation Nominal	6: race Nominal	7: gender Nominal	8: hours-per-week Nominal	9: **income** Nominal
1	20-30	Private	11th	Never-married	Machine-op-inspct	Black	Male	20-40	(=50K
2	30-40	Private	HS-grad	Married-civ-spouse	Farming-fishing	White	Male	40-60	(=50K
3	20-30	Local-gov	Assoc-acdm	Married-civ-spouse	Protective-serv	White	Male	20-40	)50K
4	40-50	Private	Some-college	Married-civ-spouse	Machine-op-inspct	Black	Male	20-40	)50K
5	30-40	Private	10th	Never-married	Other-service	White	Male	20-40	(=50K
6	60-70	Self-emp-not-inc	Prof-school	Married-civ-spouse	Prof-specialty	White	Male	20-40	)50K
7	20-30	Private	Some-college	Never-married	Other-service	White	Female	20-40	(=50K
8	50-60	Private	7th-8th	Married-civ-spouse	Craft-repair	White	Male	0-20	(=50K
9	60-70	Private	HS-grad	Married-civ-spouse	Machine-op-inspct	White	Male	20-40	)50K
10	30-40	Federal-gov	Bachelors	Married-civ-spouse	Adm-clerical	White	Male	20-40	(=50K
11	20-30	Private	HS-grad	Never-married	Adm-clerical	White	Female	20-40	(=50K
12	40-50	Private	HS-grad	Married-civ-spouse	Machine-op-inspct	White	Male	40-60	)50K
13	40-50	Private	Masters	Married-civ-spouse	Exec-managerial	White	Male	40-60	)50K
14	0-20	State-gov	Some-college	Never-married	Other-service	White	Male	20-40	(=50K
15	40-50	Private	HS-grad	Married-civ-spouse	Adm-clerical	White	Female	20-40	(=50K
16	30-40	Private	HS-grad	Widowed	Machine-op-inspct	White	Female	0-20	(=50K
17	30-40	Private	Doctorate	Married-civ-spouse	Prof-specialty	Asian...	Male	40-60	)50K
18	30-40	Private	Bachelors	Married-civ-spouse	Tech-support	White	Male	40-60	)50K
19	30-40	Private	Some-college	Never-married	Other-service	Black	Female	20-40	(=50K
20	20-30	Private	Bachelors	Never-married	Prof-specialty	White	Male	40-60	(=50K
21	20-30	Private	Bachelors	Married-civ-spouse	Prof-specialty	White	Male	20-40	(=50K
22	40-50	Self-emp-not-inc	HS-grad	Married-civ-spouse	Craft-repair	White	Male	80-100	)50K
23	20-30	Private	HS-grad	Never-married	Adm-clerical	White	Male	0-20	(=50K
24	20-30	Private	HS-grad	Separated	Machine-op-inspct	Black	Male	40-60	(=50K
25	50-60	Private	HS-grad	Married-civ-spouse	Craft-repair	White	Male	20-40	(=50K
26	30-40	Self-emp-not-inc	Some-college	Never-married	Prof-specialty	White	Male	40-60	(=50K
27	40-50	State-gov	Some-college	Married-civ-spouse	Exec-managerial	Black	Male	20-40	)50K
28	50-60	Self-emp-not-inc	11th	Widowed	Other-service	White	Female	40-60	(=50K
29	20-30	Self-emp-not-inc	Bachelors	Never-married	Sales	White	Male	40-60	(=50K
30	20-30	Local-gov	Some-college	Married-civ-spouse	Protective-serv	White	Male	20-40	(=50K
31	20-30	Private	HS-grad	Divorced	Exec-managerial	White	Female	20-40	(=50K
32	30-40	Local-gov	Bachelors	Married-civ-spouse	Prof-specialty	White	Male	20-40	)50K
33	20-30	Private	5th-6th	Never-married	Priv-house-serv	White	Male	40-60	(=50K
34	0-20	Private	10th	Never-married	Machine-op-inspct	White	Male	20-40	(=50K
35	0-20	Private	HS-grad	Never-married	Craft-repair	White	Male	20-40	(=50K
36	60-70	Private	Masters	Married-civ-spouse	Prof-specialty	White	Male	40-60	)50K
37	40-50	Self-emp-inc	Assoc-voc	Married-civ-spouse	Sales	White	Male	40-60	)50K
38	30-40	Private	HS-grad	Married-civ-spouse	Farming-fishing	White	Male	20-40	(=50K
39	20-30	Private	11th	Married-civ-spouse	Other-service	White	Male	20-40	(=50K
40	0-20	State-gov	Some-college	Never-married	Farming-fishing	White	Male	20-40	(=50K

1. 使用weka回答以下問題：

(a)請嘗試著修改 adult.csv 的欄位與上圖相同，並轉換成 arff 檔使 weka可以執行 Association Rule，請說明使用的方法以及解釋原來的 檔案不能執行的原因？(10%)

(b) 請將 numRules 設成 5 和 10，其各別執行後的 Minimum support 為 何，請比較兩者並說明造成其差異的原因。(15%)

(c) 將 numRule 設成 10，列出前 5 條 rule(15%)

(d) 如何在 Associator output 產生 Itemset，請截圖說明並附上 Itemset 結果。(15%)

2. 使用python回答以下問題：

(e) 使用已修改過的 adult.csv 檔，使用 Apriori 演算法進行分析,設定 confidence = 0.9、minimum support = 0.2，過程中對所有重要程式步驟進行截圖並加以說明，越詳盡越好。(15%)

(f) 產生與 (c) 小題一樣的結果，列出前五條best rules，截圖並加以說明(15%)

- 繳交期限: 4/1(三) 中午 12:00

- 請轉檔為 PDF 格式,檔名為:ECT_HW2_學號_版本.pdf,並同時附上

- python 的 ipynb.檔，命名格式同上。

- 上傳至 LMS 作業區,遲交一天扣該次作業成績 5%

- 補交請上傳至 HW2 補交區,僅開放到 4/8(三) 中午 12:00