Association Analysis

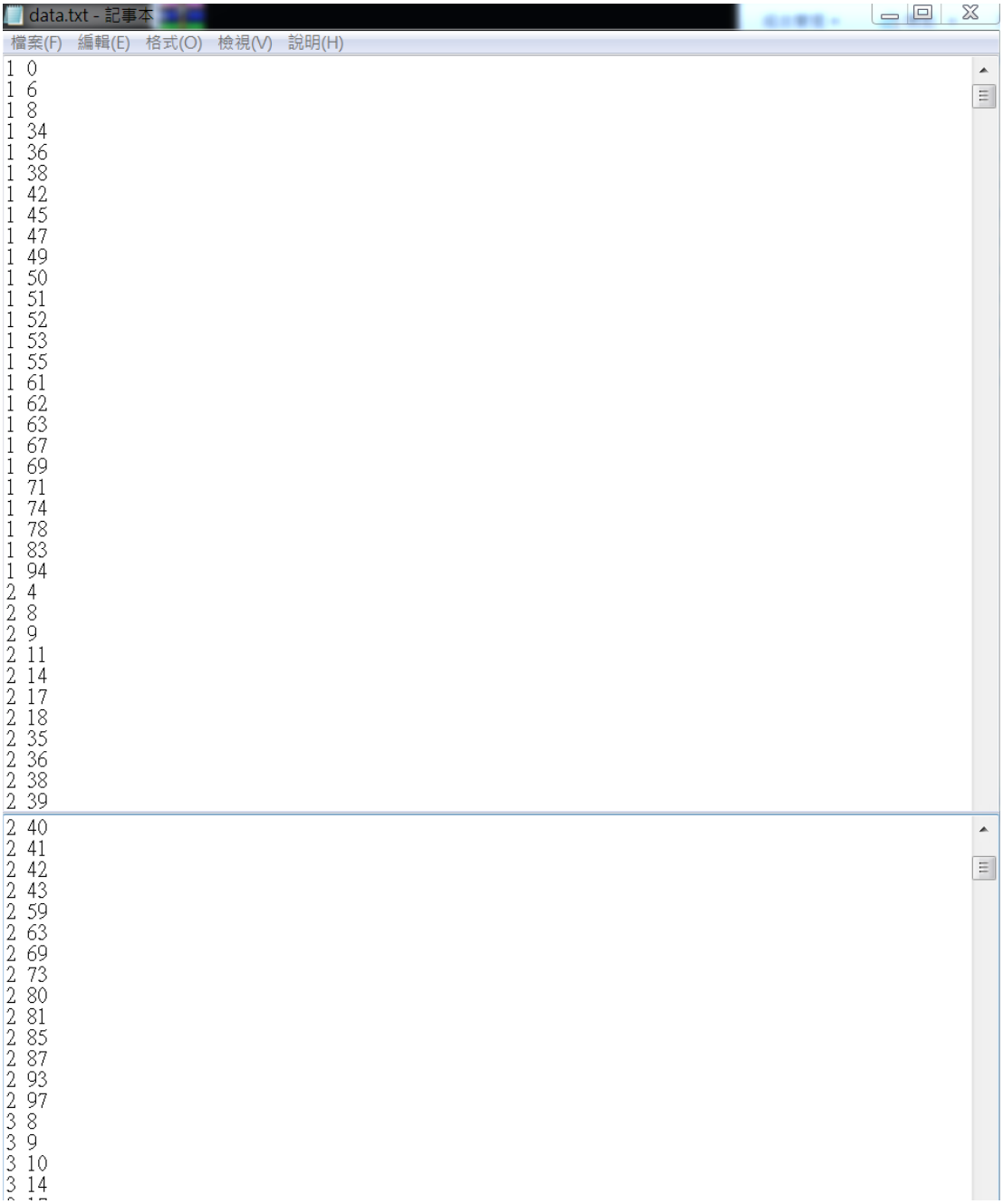
E94056178

Dataset: data.txt (Use IBM Quest Synthetic Data Generator)

Apriori code : ” Apriori.py” or “Apriori.ipynb” (Python3 執行產生結果)

FP-growth code : “FP-growth.py” or “FP-growth.ipynb” (Python3 執行產生結果)

Dataset(data.txt)



.....未列完

Apriori result (min support = 0.2 confidence = 0.4)

```
Rule:frozenset({'38', '63'})-->frozenset({'8'}) conf:0.4651162790697675 support:0.2
Rule:frozenset({'8', '63'})-->frozenset({'38'}) conf:0.7142857142857143 support:0.2
Rule:frozenset({'8', '38'})-->frozenset({'63'}) conf:0.6451612903225807 support:0.2
Rule:frozenset({'8'})-->frozenset({'38', '63'}) conf:0.4444444444444445 support:0.2
Rule:frozenset({'38', '63'})-->frozenset({'36'}) conf:0.627906976744186 support:0.27
Rule:frozenset({'36', '63'})-->frozenset({'38'}) conf:0.675 support:0.27
Rule:frozenset({'36', '38'})-->frozenset({'63'}) conf:0.7105263157894737 support:0.27
Rule:frozenset({'63'})-->frozenset({'36', '38'}) conf:0.435483870967742 support:0.27
Rule:frozenset({'69', '38'})-->frozenset({'36'}) conf:0.6363636363636364 support:0.21
Rule:frozenset({'36', '38'})-->frozenset({'69'}) conf:0.5526315789473684 support:0.21
Rule:frozenset({'36', '69'})-->frozenset({'38'}) conf:0.6176470588235293 support:0.21
Rule:frozenset({'69', '63'})-->frozenset({'36'}) conf:0.7352941176470588 support:0.25
Rule:frozenset({'36', '63'})-->frozenset({'69'}) conf:0.625 support:0.25
Rule:frozenset({'36', '69'})-->frozenset({'63'}) conf:0.7352941176470588 support:0.25
Rule:frozenset({'63'})-->frozenset({'36', '69'}) conf:0.40322580645161293 support:0.25
Rule:frozenset({'69'})-->frozenset({'36', '63'}) conf:0.45454545454545453 support:0.25
Rule:frozenset({'36'})-->frozenset({'69', '63'}) conf:0.4385964912280702 support:0.25
Rule:frozenset({'38', '63'})-->frozenset({'69'}) conf:0.4651162790697675 support:0.2
Rule:frozenset({'69', '63'})-->frozenset({'38'}) conf:0.5882352941176471 support:0.2
Rule:frozenset({'69', '38'})-->frozenset({'63'}) conf:0.6060606060606061 support:0.2
Rule:frozenset({'17', '38'})-->frozenset({'36'}) conf:0.7 support:0.21
Rule:frozenset({'36', '38'})-->frozenset({'17'}) conf:0.5526315789473684 support:0.21
Rule:frozenset({'36', '17'})-->frozenset({'38'}) conf:0.6774193548387096 support:0.21
Rule:frozenset({'38', '85'})-->frozenset({'36'}) conf:0.65625 support:0.21
Rule:frozenset({'36', '85'})-->frozenset({'38'}) conf:0.7777777777777777 support:0.21
Rule:frozenset({'36', '38'})-->frozenset({'85'}) conf:0.5526315789473684 support:0.21
Rule:frozenset({'85'})-->frozenset({'36', '38'}) conf:0.4883720930232558 support:0.21
Rule:frozenset({'87', '38'})-->frozenset({'36'}) conf:0.5714285714285714 support:0.24
Rule:frozenset({'36', '87'})-->frozenset({'38'}) conf:0.7272727272727272 support:0.24
Rule:frozenset({'36', '38'})-->frozenset({'87'}) conf:0.631578947368421 support:0.24
Rule:frozenset({'87'})-->frozenset({'36', '38'}) conf:0.41379310344827586 support:0.24
Rule:frozenset({'17', '63'})-->frozenset({'36'}) conf:0.6896551724137931 support:0.2
Rule:frozenset({'36', '63'})-->frozenset({'17'}) conf:0.5 support:0.2
Rule:frozenset({'36', '17'})-->frozenset({'63'}) conf:0.6451612903225807 support:0.2
Rule:frozenset({'43', '63'})-->frozenset({'36'}) conf:0.7499999999999999 support:0.21
Rule:frozenset({'36', '63'})-->frozenset({'43'}) conf:0.5249999999999999 support:0.21
Rule:frozenset({'36', '43'})-->frozenset({'63'}) conf:0.7499999999999999 support:0.21
Rule:frozenset({'87', '63'})-->frozenset({'36'}) conf:0.6875 support:0.22
Rule:frozenset({'36', '87'})-->frozenset({'63'}) conf:0.6666666666666666 support:0.22
Rule:frozenset({'36', '63'})-->frozenset({'87'}) conf:0.5499999999999999 support:0.22
Rule:frozenset({'43', '69'})-->frozenset({'36'}) conf:0.7096774193548387 support:0.22
```

.....未列完

FP-growth result

```
frequency items:
{'36'}
{'38', '36'}
{'63', '38', '36'}
{'38', '36', '87'}
{'63', '38', '36', '87'}
{'63', '36'}
{'87', '36'}
{'63', '87', '36'}
{'38'}
{'63'}
{'63', '38'}
{'69'}
{'69', '36'}
{'69', '36', '38'}
{'63', '69', '36', '38'}
{'63', '69', '36'}
{'69', '36', '87'}
{'63', '69', '36', '87'}
{'69', '38'}
{'69', '36', '38'}
{'63', '69', '36', '38'}
{'36', '38', '63', '87', '69'}
{'69', '87', '36', '38'}
{'63', '69', '38'}
{'63', '69', '87', '38'}
{'69', '87', '38'}
{'63', '69'}
{'69', '87'}
{'63', '69', '87'}
{'87'}
{'87', '38'}
{'63', '87', '38'}
{'63', '87'}
```

.....未列完

¨ Use association analysis tools (沒使用 WEKA 用 python 的 apyori 套件比較)

相關 code: “Apriori_tool.py” or “Apriori_tool.ipynb”

```
Rule: 17 38  -> 36 conf:0.7 support :0.21 lift: 1.2280701754385965
Rule: 17 63  -> 36 conf:0.6451612903225807 support :0.2 lift: 1.040582726326743
Rule: 17 87  -> 36 conf:0.6774193548387096 support :0.21 lift: 1.167964404894327
Rule: 17 38  -> 63 conf:0.7 support :0.21 lift: 1.129032258064516
Rule: 17 38  -> 87 conf:0.7 support :0.21 lift: 1.206896551724138
Rule: 28 3  -> 38 conf:0.7499999999999999 support :0.21 lift: 1.0869565217391304
Rule: 28 38  -> 48 conf:0.5555555555555556 support :0.2 lift: 1.2345679012345678
Rule: 28 38  -> 63 conf:0.7692307692307693 support :0.2 lift: 1.1148272017837237
Rule: 28 38  -> 69 conf:0.6111111111111112 support :0.22 lift: 1.1111111111111112
Rule: 28 38  -> 87 conf:0.6666666666666666 support :0.24 lift: 1.1494252873563218
Rule: 28 87  -> 69 conf:0.6666666666666667 support :0.2 lift: 1.149425287356322
Rule: 3 38  -> 87 conf:0.7096774193548387 support :0.22 lift: 1.2235817575083427
Rule: 63 38  -> 36 conf:0.7105263157894737 support :0.27 lift: 1.1460101867572157
Rule: 69 38  -> 36 conf:0.5526315789473684 support :0.21 lift: 1.0047846889952152
Rule: 38 36  -> 85 conf:0.5526315789473684 support :0.21 lift: 1.285189718482252
Rule: 38 36  -> 87 conf:0.631578947368421 support :0.24 lift: 1.0889292196007259
Rule: 63 36  -> 43 conf:0.7499999999999999 support :0.21 lift: 1.2096774193548385
Rule: 69 36  -> 43 conf:0.7857142857142857 support :0.22 lift: 1.4285714285714284
Rule: 69 63  -> 36 conf:0.625 support :0.25 lift: 1.1363636363636362
Rule: 63 87  -> 36 conf:0.6666666666666666 support :0.22 lift: 1.075268817204301
Rule: 38 63  -> 43 conf:0.7692307692307693 support :0.2 lift: 1.240694789081886
Rule: 8 38  -> 63 conf:0.4651162790697675 support :0.2 lift: 1.03359173126615
Rule: 38 63  -> 85 conf:0.4651162790697675 support :0.2 lift: 1.081665765278529
Rule: 38 63  -> 87 conf:0.71875 support :0.23 lift: 1.0416666666666667
Rule: 69 38  -> 87 conf:0.696969696969697 support :0.23 lift: 1.2016718913270639
Rule: 69 63  -> 43 conf:0.8214285714285714 support :0.23 lift: 1.4935064935064932
```

.....未列完

Compare results

association analysis tool

```
Rule: 17 38  -> 36 conf:0.7 support :0.21 lift: 1.2280701754385965
Rule: 17 63  -> 36 conf:0.6451612903225807 support :0.2 lift: 1.040582726326743
Rule: 17 87  -> 36 conf:0.6774193548387096 support :0.21 lift: 1.167964404894327
Rule: 17 38  -> 63 conf:0.7 support :0.21 lift: 1.129032258064516
Rule: 17 38  -> 87 conf:0.7 support :0.21 lift: 1.206896551724138
Rule: 28 3  -> 38 conf:0.7499999999999999 support :0.21 lift: 1.0869565217391304
Rule: 28 38  -> 48 conf:0.5555555555555556 support :0.2 lift: 1.2345679012345678
Rule: 28 38  -> 63 conf:0.7692307692307693 support :0.2 lift: 1.1148272017837237
Rule: 28 38  -> 69 conf:0.6111111111111112 support :0.22 lift: 1.1111111111111112
Rule: 28 38  -> 87 conf:0.6666666666666666 support :0.24 lift: 1.1494252873563218
Rule: 28 87  -> 69 conf:0.6666666666666667 support :0.2 lift: 1.149425287356322
Rule: 3 38  -> 87 conf:0.7096774193548387 support :0.22 lift: 1.2235817575083427
Rule: 63 38  -> 36 conf:0.7105263157894737 support :0.27 lift: 1.1460101867572157
Rule: 69 38  -> 36 conf:0.5526315789473684 support :0.21 lift: 1.0047846889952152
Rule: 38 36  -> 85 conf:0.5526315789473684 support :0.21 lift: 1.285189718482252
Rule: 38 36  -> 87 conf:0.631578947368421 support :0.24 lift: 1.0889292196007259
Rule: 63 36  -> 43 conf:0.7499999999999999 support :0.21 lift: 1.2096774193548385
Rule: 69 36  -> 43 conf:0.7857142857142857 support :0.22 lift: 1.4285714285714284
Rule: 69 63  -> 36 conf:0.625 support :0.25 lift: 1.1363636363636362
```

Apriori implementation

```
Rule:frozenset({'17', '38'})-->frozenset({'63'}) conf:0.7 support:0.21
Rule:frozenset({'17', '63'})-->frozenset({'36'}) conf:0.6896551724137931 support:0.2
Rule:frozenset({'17', '38'})-->frozenset({'36'}) conf:0.7 support:0.21
Rule:frozenset({'38', '28'})-->frozenset({'69'}) conf:0.6111111111111112 support:0.22
Rule:frozenset({'87', '28'})-->frozenset({'69'}) conf:0.6666666666666667 support:0.2
Rule:frozenset({'38', '28'})-->frozenset({'63'}) conf:0.5555555555555556 support:0.2
```

「皆節錄部分結果」同 rule 的 confidence 與 support 一致

Find and Answer

對 data.txt 做 APriori 關聯分析(自己實作),不同 support 與 confidence 結果如下:

Rules with High support, high confidence

```
Rule:frozenset({'36'})-->frozenset({'63'}) conf:0.7017543859649124 support:0.4
Rule:frozenset({'87'})-->frozenset({'38'}) conf:0.7241379310344828 support:0.42
```

Rules with High support(0.4), Low confidence(0.2)

```
Rule:frozenset({'63'})-->frozenset({'36'}) conf:0.6451612903225807 support:0.4
Rule:frozenset({'36'})-->frozenset({'63'}) conf:0.7017543859649124 support:0.4
Rule:frozenset({'38'})-->frozenset({'63'}) conf:0.6231884057971014 support:0.43
Rule:frozenset({'63'})-->frozenset({'38'}) conf:0.6935483870967741 support:0.43
Rule:frozenset({'38'})-->frozenset({'87'}) conf:0.6086956521739131 support:0.42
Rule:frozenset({'87'})-->frozenset({'38'}) conf:0.7241379310344828 support:0.42
```

Rules with Low support(0.2), high confidence(0.8)

```
Rule:frozenset({'43', '63'})-->frozenset({'69'}) conf:0.8214285714285714 support:0.23
Rule:frozenset({'87', '3'})-->frozenset({'38'}) conf:0.88 support:0.22
Rule:frozenset({'87', '28'})-->frozenset({'38'}) conf:0.8 support:0.24
Rule:frozenset({'28', '48'})-->frozenset({'38'}) conf:0.8 support:0.2
```

Rules with Low support(0.2), Low confidence(0.2)

```
Rule:frozenset({'38', '63'})-->frozenset({'8'}) conf:0.4651162790697675 support:0.2
Rule:frozenset({'8', '63'})-->frozenset({'38'}) conf:0.7142857142857143 support:0.2
Rule:frozenset({'8', '38'})-->frozenset({'63'}) conf:0.6451612903225807 support:0.2
Rule:frozenset({'63'})-->frozenset({'8', '38'}) conf:0.32258064516129037 support:0.2
Rule:frozenset({'38'})-->frozenset({'8', '63'}) conf:0.2898550724637682 support:0.2
Rule:frozenset({'8'})-->frozenset({'38', '63'}) conf:0.4444444444444445 support:0.2
Rule:frozenset({'38', '63'})-->frozenset({'36'}) conf:0.627906976744186 support:0.27
Rule:frozenset({'36', '63'})-->frozenset({'38'}) conf:0.675 support:0.27
Rule:frozenset({'36', '38'})-->frozenset({'63'}) conf:0.7105263157894737 support:0.27
Rule:frozenset({'63'})-->frozenset({'36', '38'}) conf:0.435483870967742 support:0.27
Rule:frozenset({'38'})-->frozenset({'36', '63'}) conf:0.39130434782608703 support:0.27
Rule:frozenset({'36'})-->frozenset({'38', '63'}) conf:0.4736842105263159 support:0.27
Rule:frozenset({'69', '38'})-->frozenset({'36'}) conf:0.6363636363636364 support:0.21
Rule:frozenset({'36', '38'})-->frozenset({'69'}) conf:0.5526315789473684 support:0.21
Rule:frozenset({'36', '69'})-->frozenset({'38'}) conf:0.6176470588235293 support:0.21
Rule:frozenset({'38'})-->frozenset({'36', '69'}) conf:0.30434782608695654 support:0.21
Rule:frozenset({'69'})-->frozenset({'36', '38'}) conf:0.3818181818181818 support:0.21
Rule:frozenset({'36'})-->frozenset({'69', '38'}) conf:0.368421052631579 support:0.21
Rule:frozenset({'69', '63'})-->frozenset({'36'}) conf:0.7352941176470588 support:0.25
```

.....共 162 列

此資料均為純數字，當作字串處理，每個元素對整體的 support 沒有太大，結果的 support 代表該 rule 整體出現在 data 的頻率，confidence 代表該 rule 前面關聯後面的可能性多大。以市場交易為例：高 support 高 confidence 可能代表事物性質相近，一起購買的機率高，交易比重也高；高 support 低 confidence 交易比重高，但是一起購買機率不高；低 support 高 confidence 雖然事物一起買機率高，但是在市場交易比重占太少；低 support 低 confidence 則代表兩事物關聯性低。