

Data Mining 2

Module 4 - 2020/2021

Name _____ Surname _____ ID: _____

Test id. AUTO

Answers

Q1: The Apriori principle, namely $S_1 \subseteq S_2 \rightarrow support(S_1) \geq support(S_2)$

Q2: Pruned: $\{A\} \rightarrow \{C\} \rightarrow \{A\}$

Frequent: $\{A\} \rightarrow \{C\} \rightarrow \{E\}$ and $\{C\} \rightarrow \{C\} \rightarrow \{C\}$

Q3: $gap \leq 3$ is better, because it does not consider useless "before-after lockdown" occurrences like $\langle 0, 7 \rangle$ or $\langle 1, 7 \rangle$ (and others), while $gap \geq 3$ does not consider useful occurrences like $\langle 5, 6 \rangle$ or $\langle 6, 7 \rangle$, namely those during the lockdown.

Q4: 2, 4

Q5: 4, 5

Q6: $C(P_1, P_2, P_4), C(P_3)$

Q7: $Profit(Partition1) = 1.75$

$Profit(Partition2) = 1.08$

List id questions: [1, 2, 3, 4, 5, 6, 7]