Data Mining 2

CAT 4 - 2019/2020

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*** $< 0, 7 >$ ***or*** $< 1, 7 >$ ***(and others), while*** $gap \geq 3$ ***does not consider useful occurrences like*** $< 5, 6 >$ ***or*** $< 6, 7 >$***, namely those during the lockdown.***

Q4: 3, 4

Q5: $C(P_1, P_2, P_4), C(P_3)$

Q6: $Profit(Partiton1) = 1.75$
$Profit(Partition2) = 1.08$

Q7: 1

Q8: B, E

Q9: $LOF(p) = \frac{\sum_{o \in knn(p)} \frac{lrd(p)}{lrd(o)}}{|knn(p)|}$
where $lrd(p) = \frac{1}{\frac{\sum_{o \in knn(p)} rd(p,o)}{|knn(p)|}}$
where $rd(p, o) = max\{k - distance(o), dist(p, o)\}$

List id questions: [1, 2, 3, 4, 5, 6, 7, 8, 9]