

Examen Knowledge management and business intelligence

6 points from Assignment 1 and 2

Exam (3 hours, professor Vanthienen)

10 points from 15 MC questions (guess correction)

4 points on open questions (4 questions, each on 1 point)

In general: do not study formulas by hard, he gives them when you would need them. You need to be able to apply them. Level of the exam was okay. Some more difficult MCs, asked question from almost every chapter. Did not ask anything from RapidMiner (or interpreting its results), nor of the AE and Colruyt guest lectures.

Multiple Choice Questions

1. Indicate correct statement topic preprocessing
2. Indicate correct statement topic clustering
3. Definitions Accuracy, sensitivity, specificity
4. Combi Roc and Accuracy, sensitivity, specificity (True-positive rate, positive rate)
→ Which one is correct
 - a. $FP\text{-rate} = 1\text{-specificity}$ I thought this one, calculated it with examples for TP, TN, FP, FN and thought of this formula for FPrate: $FP/FP+TN$ (could be wrong)
 - b. $FP\text{-rate} = 1\text{-sensitivity}$
5. Which statement data warehousing <-> databases is in general correct?
 - a. DWs updated more frequently than operational dbs
 - a. DW does not need to be normalized
I was doubting between A and B!
 - b. Star scheme is normalized
 - c. ...
6. Forward chaining (cf. typical exam questions)
7. Calculate an entropy (formula also given) of 1000 instances, 500 in one class, 500 in other class
 . **1.0** $=(\log(1/2)/\log(2))$
 - a. 0.33
 - b. 0.5
 - c. 0
8. Four statements on visualization, which one is part of the visualization rules (harder question, answer possibilities were something like this)
 - a. Data ink = 0 means you need all graphs
 - a. Numbers in graph always have to be proportional to quantities on axis
 - b. Money needs to be standardized and deflated or nominal on axis
 - c. Dimensions should always be larger than what's in the graph
9. Difficult question CH7: Decision models and mining
 - a. P-MinD
 - b. ...
10. What hit type if you will always have the same answer? (CH14)
 - a. **any hit**
 - b. first hit
 - c. unique hit
 - d. multiple hit
11. Very nice question fuzzy reasoning: What is $A \cup B$ (formula given)

Set1 - (1, 0.2); (2, 0.3); (3, 0.5); (4, 0.4); (5, 0.1)

Set2 - (1, 0.3); (2, 0.4); (3, 0.5); (4, 0.3); (5, 0.2)

 - a. (1, 0.2); (2, 0.3); (3, 0.5); (4, 0.3); (5, 0.1)
 - b. (1, 0.5); (2, 0.7); (3, 1.0); (3, 0.7); 5(0.3)

c. ...

Open questions

1. Apriori
2. Calculate information gain (goodness of the split)
3. 12 MAKES, discuss if low volatility and service based has a codification or personalization approach
4. What is the name of the redundancy rule called that... + give an example
→ Subsumption

