

DM1 – zkouškové otázky:

1. Vysvětlete pojem Data mining a popište hlavní rozdíly oproti statistické analýze.
2. Podrobně popište metodologii CRISP-DM.
3. Podrobně popište metodologii SEMMA.
4. Vysvětlete rozdíl mezi prediktivním a deskriptivním modelováním, uveďte příklady metod/modelů, vysvětlete pojmy supervised a unsupervised learning.
5. Vysvětlete pojem „libref“ v systému SAS.
6. Uveďte 5 možností jak importovat data do systému SAS, stručně popište (ne SAS kódy).
7. Jaký je rozdíl mezi formátem a informátem v systému SAS?
8. Relační databáze.
9. Vysvětlete pojmy ODS, DWH, Datamart, ETL, Meta data.
10. Datové modely - uveďte a popište rozdíly.
11. Datová kostka, OLAP operace, architektura OLAP serverů.
12. Popište a vysvětlete (ne SAS kódy) pojmy Inner/left/right/full/cross join v rámci proc SQL v SAS.
13. Čištění dat.
14. Dummy proměnné.
15. WOE – popište výpočet a interpretaci.
16. Popište rozdíly mezi data stepem a proc SQL v systému SAS.
17. Popište first. a last. proměnné v SAS data step.
18. Uveďte způsoby spojování tabulek pomocí SAS data stepu – popište (ne SAS kódy).
19. Explorační analýza.
20. Lineární regrese v SAS (ne SAS kódy), Mallowovo Cp, AIC, SBC, vlivná pozorování.
21. Logistická regrese v SAS (ne SAS kódy), model, odhad parametrů.
22. Konkordantní, diskordantní a shodné páry.
23. Rozhodovací stromy.
24. Neuronové sítě.
25. Uveďte statistiky/indexy používané pro evaluaci modelů, stručně je popište.
26. Lorenzova a CAP křivka – definujte a vysvětlete rozdíly.
27. Definujte Lift, QLift, LR.
28. Informační hodnota: obecně + za předpokladu normality skóre.
29. ROC.
30. Uveďte možnosti exportu dat ze systému SAS, stručně popište (ne SAS kódy).