

# **Témata projektů z M8DM1**

## **Datový soubor Adult**

1. Predikujte, zda příjem dané osoby překročí 50 000 USD. Data nejprve rozdělte na trénovací a testovací a poté váš model vyzkoušejte na testovacích datech.

## **Datový soubor Adult2**

2. Predikujte, zda příjem dané osoby překročí 50 000 USD. Data nejprve rozdělte na trénovací a testovací a poté váš model vyzkoušejte na testovacích datech.

## **Datový soubor Airline**

3. Které letecké společnosti jsou nejbezpečnější? A které naopak nejvíce nebezpečné?

## **Datový soubor Auto**

4. Predikujte spotřebu auta v závislosti na zadaných veličinách. Na kterých z nich nejvíce závisí? Data nejprve rozdělte na trénovací a testovací a poté váš model vyzkoušejte na testovacích datech.
5. Která auta jsou si nejvíce podobná? Která se naopak nejvíce liší? Dají se mezi auty nalézt nějaké charakteristické skupiny aut?

## **Datový soubor Bank**

6. Najděte faktory ovlivňující to, zda klient zareaguje na telemarketing a založí si termínovaný vklad.

## **Datový soubor Boston**

7. Predikujte cenu domu (medián) v Bostonu na základě socio-ekonomických ukazatelů. Na jakých attributech závisí? Data nejprve rozdělte na trénovací a testovací a poté váš model vyzkoušejte na testovacích datech.
8. Predikujte koncentraci oxidů dusíku v různých částech Bostonu na základě socio-ekonomických ukazatelů. Na jakých attributech závisí? Data nejprve rozdělte na trénovací a testovací a poté váš model vyzkoušejte na testovacích datech.

## **Datový soubor Cars93**

9. Dá se pro každého výrobce najít typický vůz, který vyrábí? Jak se různí výrobci liší?
10. Která auta jsou si nejvíce podobná? Která se naopak nejvíce liší? Dají se mezi auty nalézt nějaké charakteristické skupiny aut?

## **Datový soubor Fertility**

11. Identifikujte rizikové faktory neplodnosti u mužů.

## **Datový soubor Flags**

12. Na základě údajů o vlajce dané země predikujte nejrozšířenější náboženství v zemi.
13. Vlajky kterých zemí jsou si nejvíce podobné? Dají se nalézt skupiny zemí, které mají podobné vlajky?

## **Datový soubor German**

14. Najděte faktory ovlivňující to, zda bude klient splácet půjčku.

## **Datový soubor Health**

15. Porovnejte státy USA z hlediska zdraví. Které státy jsou na tom podobně? Ve kterém státě je situace nejlepší/nejhorsí? Souvisí geografická poloha státu se zdravím?

## **Datový soubor Heart**

16. Na základě údajů o pacientech rozhodněte, jestli daná osoba trpí onemocněním srdce. Data nejprve rozdělte na trénovací a testovací a poté váš model vyzkoušejte na testovacích datech.

## **Datový soubor Internet**

17. Jaká je situace v ČR z hlediska využívání internetu? Jaké služby nejčastěji lidé využívají? Které služby jsou používány současně?
18. Popište skladbu lidí a charakteristické skupiny, které využívají internet a jednotlivé jeho služby. Můžeme mezi nimi pozorovat nějaké rozdíly?
19. Liší se využívání internetových služeb mezi muži a ženami? Mezi bohatšími a chudšími? Napříč věkovými kategoriemi? Podle jiných aspektů?

## **Datový soubor Jobs**

20. Zkoumejte skladbu zaměstnanosti v jednotlivých zemích. Můžeme ji nějak charakterizovat? Najděte v datech skupiny zemí podobných ve smyslu zaměstnanosti. Čím jsou charakteristické?

## **Datový soubor Kabelovka**

21. Jaké je využívání TV a různých technologií příjmu signálu v ČR? Liší se nějakým způsobem podle socio-demografických charakteristik?
22. Jak vypadá častý (pravidelný) divák TV? Jak se liší od běžného diváka?
23. Nová televizní stanice chce vstoupit na český trh. Jak by se měla profilovat, aby odpovídala častým a typickým potřebám české populace?
24. Analyzujte uvedené výroky. Dají se vypožorovat nějaké vztahy mezi jednotlivými výroky? Jak souvisí se sledováním TV?
25. Analyzujte uvedené výroky v závislosti na socio-demografických charakteristikách.

## **Datový soubor Mazlicci**

26. Jaká je situace v ČR z pohledu domácích mazlíčků? Kde pro ně lidé nejčastěji nakupují krmení? Dají se nalézt nějaké zákonitosti, ve kterých obchodech lidé nakupují?
27. Kdo je typický chovatel? Liší se nějak chovatelé koček od chovatelů psů?

## **Datový soubor Mushroom**

28. Na základě údajů o houbách rozhodněte, jestli je daná houba jedovatá nebo jedlá. Data nejprve rozdělte na trénovací a testovací a poté váš model vyzkoušejte na testovacích datech.

## **Datový soubor Postoje**

29. Analyzujte výroky = charakteristiky dosavadního života. Jaké výroky jsou si podobné a jaké jsou naopak opačné? Lze výroky zredukovat do menšího počtu tvrzení?
30. Analyzujte výroky = charakteristiky dosavadního života. Dají se vypožorovat nějaké vztahy mezi jednotlivými výroky?
31. Najděte v populaci skupiny lidí s podobnými postoji podle analyzovaných výroků.
32. Lze vysledovat nějaký vztah mezi výroky a demografickými charakteristikami?

## **Datový soubor Protein**

33. Najděte v datech skupiny zemí podobných ve smyslu spotřeby bílkovin. Která země je na tom nejlépe? Která naopak nejhůře? Na čem spotřeba bílkovin závisí?

## **Datový soubor Records**

34. Najděte v datech skupiny zemí s podobnými atletickými rekordy. Která země má nejhodnotnější rekordy? A která nejméně hodnotné?

## **Datový soubor Saheart**

35. Odhalte rizikové faktory pro onemocnění srdce.
36. Lze v datech odhalit skupiny podobných lidí? Které z nich mají problémy s onemocněním srdce?

## **Datový soubor Sleep**

37. Proveďte analýzu spánku zvířat. Existuje nějaká souvislost mezi délkou spánku a ostatními atributy? Dají se mezi zvířaty nalézt nějaké skupiny podobných zvířat?

## **Datový soubor Sport**

38. Popište využívání různých druhů sportu v ČR. Dají se v populaci vypozařovat nějaké vztahy mezi provozovanými sporty?
39. Existuje rozdíl mezi sportováním bohatších a chudších občanů? Jak se vůbec bohatší lidé odlišují od běžné populace?
40. Existují v populaci nějaké typické skupiny věnující se určitým sportům?

## **Datový soubor StudentPerformance**

41. Jak se liší studenti školy Gabriel Pereira (GP) od studentů školy Mousinho da Silveira (MS)?
42. Můžeme v datech najít nějaké typické skupiny studentů? Jsou tyto skupiny stejné na obou školách?
43. Existuje souvislost mezi konzumací alkoholu a sociálními a školními atributy?

## **Datový soubor Taiwan**

44. Najděte faktory ovlivňující to, zda bude mít klient potíže splácet svoji půjčku. Data nejprve rozdělte na trénovací a testovací a poté váš model vyzkoušejte na testovacích datech.

## **Datový soubor Taiwan2**

45. Najděte faktory ovlivňující to, zda bude mít klient potíže splácet svoji půjčku. Data nejprve rozdělte na trénovací a testovací a poté váš model vyzkoušejte na testovacích datech.

## Datový soubor Tips

46. Na čem záleží velikost dýška? Jaký zákazník dá největší a jaký nejmenší dýško?

## Datový soubor USCrime

47. Porovnejte státy USA z hlediska kriminality? Které státy jsou na tom podobně? Ve kterém státě je nejméně/nejvíce bezpečno? Souvisí geografická poloha státu s jeho kriminalitou?

## Datový soubor VolnyCas

48. Jaký podíl volného času věnuje průměrný člověk různým činnostem? Liší se v trávení volného času muži od žen? Jakou roli hrají další ukazatele?
49. Liší se v nějakých aspektech lidé trávící různě svůj volný čas? Lze nějak popsat typického jedince věnujícího se různým volnočasovým aktivitám?
50. Existuje nějaký rozdíl, jak lidé tráví svůj volný čas ve všední a volný den? U kterých jedinců jsou rozdíly nejmenší (největší)?

## Datový soubor Wine

51. Na základě výsledků chemické analýzy odhalte různé typy vína. Která vína jsou si nejvíce podobná? Která se naopak nejvíce liší? Čím?
52. Na základě výsledků chemické analýzy predikujte obsah alkoholu v daném víně. Na čem závisí?

## Datový soubor WineQualityRed

53. Na základě výsledků chemické analýzy predikujte kvalitu červeného vína. Data nejprve rozdělte na trénovací a testovací a poté váš model vyzkoušejte na testovacích datech.

## Datový soubor WineQualityWhite

54. Na základě výsledků chemické analýzy predikujte kvalitu bílého vína. Data nejprve rozdělte na trénovací a testovací a poté váš model vyzkoušejte na testovacích datech.

## Datový soubor Zoo

55. Na základě vlastností zvířat najděte nejvíce podobná/nepodobná zvířata. Dají se mezi zvířaty nalézt nějaké skupiny podobných zvířat?