

Úkol C – Klasifikace dat – Alcohol consumption ZS 2016

Cíl

Cílem třetího úkolu je seznámit se s praktickým použitím klasifikačních metod na příkladě klasifikace studentů v kurzech matematiky.

Data

Data o konzumaci alkoholu u studentů na střední škole v Portugalsku. Data popisují studenty z hlediska jejich úspěšnosti v oborech matematiky a portugalština, spolu s parametry rodinného prostředí, času stráveného učením, cestováním do a ze školy, volným časem, konzumací alkoholu ve všední dny a o víkendech, apod. Konzumace alkoholu mezi studenty je velmi zajímavé téma a data studie mohou posloužit k objektivnímu posouzení jeho vlivu na studijní výsledky.

Klasifikace propadlých studentů

Pokuste se vytvořit takový na základě předložených dat klasifikátor, který by predikoval, zda student neuspěje ve studiu (propadne, proměnná G_3 je 0). Klasifikátor zhodnoťte z různých praktických i teoretických hledisek.

Požadované kroky analýzy

- Vyberte alespoň dva klasifikátory vhodné pro tento typ dat. Podle čeho budete vybírat? [2 b]
- Na vhodné podmnožině dat vybrané klasifikátory natrénujte. Jaké úspěšnosti dosahují? [2 b]
- Podle jakých příznaků se klasifikátor rozhoduje? Dává to smysl? Lze na základě vaší analýzy omezit počet měřených příznaků při zachování stejné úspěšnosti klasifikace? [4 b]
- Jakou úspěšnost klasifikace očekáváte v hypotetickém reálném nasazení vašeho klasifikátoru, tj. v případě nově zapsaného studenta? [4 b]
- Jaká je pravděpodobnost na základě dat, že nově zapsaný student propadne? [2 b]
- Jaká je pravděpodobnost, že nově zapsaný student bude klasifikován jako neúspěšný student? Jaká bude naproti tomu pravděpodobnost, že nově zapsaný student bude klasifikován jako úspěšný student. Výsledky diskutujte. [4 b]
- Má ve vašem případě na přesnost klasifikace vliv to, zda je trénovací (testovací) množina vyvážená? Pokud ano, jaký? Vyvažovali jste trénovací (testovací) množinu? Pokud ano, proč a jak? Pokud ne, proč ne? [2 b]

Výsledky upravte do formy zprávy, která bude obsahovat stručný **úvod**, popis metod, které jste použili, v sekci **metody**, výsledky jejich aplikace na data v sekci **výsledky** a závěry, které jste zjistili interpretací výsledků v sekci **závěr**. Maximální délka zprávy je 3 stránky. Zprávu ve formátu pdf odevzdejte pomocí UploadSystemu.

school	škola, kterou student navštěvuje ('GP' - Gabriel Pereira nebo 'MS' - Mousinho da Silveira)
sex	studentovo pohlaví ('F' – female/žena nebo 'M' – male/muž)
age	studentův věk (od 15 do 22)
address	typ studentova bydliště ('U' – urban/městské nebo 'R' – rural/venkovské)
famsize	velikost rodiny ('LE3' - <=3 nebo 'GT3' - > 3)
Pstatus	soužití rodičů ('T' - living together/žijí spolu nebo 'A' – apart/rozvedeni)
Medu	vzdělání matky (0 - žádné, 1 – první stupeň ZŠ, 2 druhý stupeň ZŠ, 3 SŠ or 4 VŠ)
Fedu	vzdělání otce
Mjob	zaměstnání matky
Fjob	zaměstnání otce
reason	důvod pro studium na dané škole
guardian	opatrovník
traveltime	doba cesty do školy (1 - <15 min., 2 - 15 až 30 min., 3 - 30 min. to 1 hod, nebo 4 - >1 hod)
studytime	týdenní doba studia(1 - <2 hod, 2 - 2 to 5 hod, 3 - 5 až 10 hod, nebo 4 - >10 hod)
failures	počet předchozích nedokončení ročníku
schoolsup	podpora ze strany školy (yes nebo no)
famsup	vzdělávací podpora v rodině (yes nebo no)
paid	doučování (yes nebo no)
activities	mimoškolní aktivity (yes nebo no)
nursery	navštěvoval školku (yes nebo no)
higher	chce pokračovat na VŠ (yes nebo no)
internet	přístup k internetu doma (yes nebo no)
romantic	ve vztahu (yes nebo no)
famrel	kvalita domácích vztahů (od 1 - very bad do 5 – excellent)
freetime	volný čas po škole (od 1 - very low do 5 - very high)
goout	chodí ven s kamarády (od 1 - very low do 5 - very high)
Dalc	konzumace alkoholu ve všední dny (od 1 - very low do 5 - very high)
Walc	konzumace alkoholu o víkendech (od 1 - very low do 5 - very high)
health	zdravotní stav (od 1 - very bad do 5 - very good)
absences	počet absencí (od 0 do 93)
G1	Známka v prvním pololetí (od 0 do 20)
G2	Známka v druhém pololetí (od 0 do 20)
G3	Závěrečná známka (od 0 do 20)

Tabulka 1: Popis proměnných z studie konzumace alkoholu mezi studenty