

## Úkol C – Klasifikace dat – jaterní testy ZS 2018

### Cíl

Cílem třetího úkolu je seznámit se s praktickým použitím klasifikačních metod na příkladu klasifikace pacientů s cirhózou jater.

### Data

Dataset Indian Liver Patient Records obsahuje 416 záznamů pacientů s cirhózou a 167 pacientů se zdravými játry naměřených v Andhra Pradesh v Indii. Data obsahují jak základní demografické údaje, tak i biochemické markery jako je hladina bilirubinu nebo albuminu v krvi.

### Klasifikace pacientů s cirhózou jater

Na základě dat vytvořte klasifikátor, který bude predikovat, zda pacient trpí cirhózou jater nebo má játra zdravá. Klasifikátor zhodnoťte z různých praktických i teoretických hledisek.

### Požadované kroky analýzy [20b]

- 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 příchozího pacienta? [4 b]
- Jaká je pravděpodobnost na základě dat, že nový pacient bude nemocný? [2 b]
- Jaká je pravděpodobnost, že zdravý pacient bude klasifikován jako nemocný? Jaká bude naproti tomu pravděpodobnost, že nemocný pacient bude klasifikován jako zdravý? 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 5 stránek. Zprávu ve formátu *pdf* odevzdejte pomocí *UploadSystemu*.

Název parametru	Popis parametru
Age	Věk (pacientům nad 89 let je přiřazen věk 90)
Gender	Pohlaví
Total_bilirubin	Celkový bilirubin
Direct_bilirubin	Konjugovaný bilirubin
Alkaline_phosphatase	Alkalická fosfatáza
Alamine_Aminotransferase	Alaninaminotransferáza
Aspartate_Aminotransferase	Aspartátaminotransferáza
Total_protiens	Celkový sérový protein
Albumin	Albumin
Albumin_Globulin_Ration	Poměr albuminu a globulinu
Dataset	Diagnoza (1 = cirhóza jater, 2 = zdravá játra)

Tabulka 1: Výčet parametrů