

# IZU 2021/22: Zadání 3. úkolu

## Společná část zadání

Každý student má přiřazen jeden příklad, který byl zaslán e-mailem na školní účet. Pro daný příklad demonstřujete princip generování rozhodovacích klasifikačních stromů algoritmem **ID3 (Iterative Dichotomiser 3)** představeným na přednášce.

## Formát zadání příkladů

Zadání jednotlivých příkladů jsou uložena v textových souborech. Název souboru je ve formátu model-xloginXX.txt. Každé zadání je definováno následujícími seznamy v příslušném souboru (zde uvedené příklady jsou pouze ilustrativní):

### 1. Seznam atributů s výčtem hodnot.

```
attributes {
    uver : spatna nan dobra
    dluh : vysoky nizky
    ruceni : zadne adekv
    prijem : m s v
}
```

### 2. Seznam tříd, do kterých jsou objekty klasifikovány.

```
classes {
    V
    P
    N
}
```

### 3. Seznam pozorování (příkladů/objektů), z kterých se tvoří rozhodovací strom. Každý řádek odpovídá jednomu objektu:

- 1. sloupec je identifikátor objektu
- 2. sloupec je třída, do které je objekt klasifikován
- další sloupce obsahují hodnoty atributů v tom pořadí, jak jsou definovány v seznamu atributů

```
objects {
    1 V spatna vysoky zadne m
    2 V nan vysoky zadne s
    3 P nan nizky zadne s
}
```

## Řešení a formát odevzdání

Seznam objektů obsahuje vždy 12 příkladů. Vytvořte dva rozhodovací stromy podle následujících pravidel:

- 1. strom: použijte příklady 1 až 3 jako testovací sadu a příklady 4 až 12 jako trénovací sadu
- 2. strom: použijte příklady 10 až 12 jako testovací sadu a příklady 1 až 9 jako trénovací sadu
- každý strom vytvořte podle trénovací sady a validujte podle testovací sady

#### Zápis vygenerovaného stromu

- uzel je buď atribut nebo třída
- hrana je ohodnocena konkrétní hodnotou atributu
- u atributového uzlu budou navíc informace o vypočítaných ziscích (gains) pro všechny uvažované varianty; vypočtené hodnoty uvádějte na 3 desetinná místa
- u hran bude navíc informace o objektech, kterým odpovídá zvolená cesta (seznam jejich identifikátorů)

#### Zápis validace stromu

- pro každý testovací příklad vyhodnoťte správnost predikce a zapište ke stromu (čitelně a zvýrazněně) ve formátu  
id\_objektu: uzel[hodnota\_zvolene\_hrany] ->  
uzel[hodnota\_zvolene\_hrany] -> predikovaná\_třída =>  
[skutečná\_třída] FAIL|OK  
např.:  
10: příjem[s] -> dluh[nizky] -> P => [N] FAIL  
nebo:  
10: příjem[s] -> dluh[nizky] -> P => [P] OK
- poznámka: vzhledem k omezenému počtu trénovacích příkladů nelze očekávat vysokou úspěšnost klasifikace

Řešení můžete vytvořit v nástroji nebo čitelně vytvořit na papír a naskenovat. Výsledek odevzdejte elektronicky ve formátu pdf (příp. jpg).

Příklad formátu řešení je součástí zadání ([model-ukazka.pdf](#)).