

# IZU 2020/21: Zadání 3. úkolu

## Společná část zadání

Na konci období Sengoku (Japonsko přelomu 16. a 17. století), po smrti regenta Tošiie Maeda, došlo opět k bojům o vládu nad Japonskem. Šógunem se chtěl stát Iejasu Tokugawa, proti němu se postavila aliance věrných rodině Tojotomi pod vedením Micinari Išida. Tokugawa se rozhoduje, zda se utkat s protivníkem u Sekigahary. Rozhoduje se na základě předchozích znalostí a zkušeností s velikostmi armád, možnosti zrady (na své straně či straně nepřítele) a zda by byl útok očekávaný.

Vášim úkolem je vytvořit rozhodovací strom na základě těchto znalostí a odpovědět Iejasovi Tokugawovi na otázku, zda je za dané situace vhodné zaútočit. Objekty odpovídající konkrétním hodnotám vlastností jsou klasifikovány do tří možných odpovědí – Y (ano, zaútočit), N (ne, neútočit), M (uspokojivá možnost vítězství). [Poznámka: Ve skutečnosti k bitvě došlo a Tokugawa zvítězil, různé kombinace v příkladech však mohou vést k odlišným výsledkům.]

Každý student má přiřazen jeden příklad (konkrétní skupina znalostí a otázka) pod číslem 1 až 40. Pro daný příklad demonstřujete princip generování rozhodovacích stromů algoritmem **ID3 (Iterative Dichotomiser 3)** představeným na přednášce a odpovězte na otázku.

## Formát zadání příkladů

Zadání jednotlivých příkladů jsou uložena v textových souborech. Číslo příkladu odpovídá číslu v názvu souboru (např. `id3-24.txt` obsahuje zadání č. 24). 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 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
}
```

4. Seznam otázek, každý řádek odpovídá jedné otázce:

- 1. sloupec je identifikátor otázky
- další sloupce obsahují hodnoty atributů v tom pořadí, jak jsou definovány v seznamu atributů

```
questions {  
    1 dobra vysoky zadne m  
}
```

## Řešení a formát odevzdání

Řešením je vygenerovaný rozhodovací strom

- 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 4 desetinná místa
- u hran bude navíc informace o objektech, kterým odpovídá zvolená cesta (seznam jejich identifikátorů)
- součástí bude odpověď na zadanou otázku včetně cesty ve stromu (lze naznačit barevnou čarou, výpisem apod.)

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 ([id3-ukazka.pdf](#)).