

Ukázkový test k přijímacímu řízení Umělá inteligence a data science

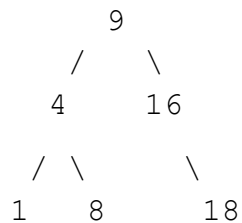
U každé otázky je uveden název okruhu a orientační informace o bodovém zastoupení tématu v testu. Test je rozdělen do tematických okruhů odpovídajících požadovaným znalostem uchazeče. V tomto ukázkovém testu nejsou plně zastoupena všechna témata z vymezených okruhů. Zároveň platí, že ne všechna konkrétní témata a typy úloh obsažené v tomto ukázkovém testu se nutně vyskytnou v reálné přijímací zkoušce.

1. Algoritmy a výpočetní složitost

1.1 Algoritmus

Indikativní bodování: 7 bodů

Mějme binární vyhledávací strom



vytvořený pomocí následující třídy v Pythonu:

```

-----
class Node:
    def __init__(self, key):
        self.key = key
        self.left = None
        self.right = None
-----
  
```

Proměnná `root` značí kořen stromu. Předpokládejme, že uzly tohoto stromu jsou instancemi výše zmíněné třídy `Node` a jsou v paměti již správně propojeny.

Uvažujme následující algoritmus:

```

-----
def insert(root, S):
    if root is None:
        return Node(S)
    else:
        if root.key == S:
            return root
  
```

```
elif root.key < S:  
    root.right = insert(root.right, S)  
else:  
    root.left = insert(root.left, S)  
return root
```

Výška stromu je definována jako délka nejdelší cesty (počet hran) od kořene k nejbzdalejšímu listu. Jaká bude výška uvažovaného binárního stromu výše po provedení následujících příkazů?

```
insert(root, 17)  
insert(root, 20)
```

Vyberte jednu z možností:

- A. Výška stromu bude 1
- B. Výška stromu bude 2
- C. Výška stromu bude 3
- D. Výška stromu bude 4
- E. Výška stromu bude 5

Správná odpověď: C

1.2 Výpočetní složitost

Indikativní bodování: 7 bodů

Jaká je výpočetní složitost (nejtěsnější horní odhad) v nejhorším případě následující funkce, jestliže n je velikost listu `arr`?

Poznámka: funkce `range(n)` vytváří sekvenci 0 až $n-1$.

```
def func(arr, target):  
    s = 0  
    for i in range(len(arr)):  
        if arr[i] == target:  
            for j in range(i + 1):  
                s = s + arr[j]  
            return s  
    return s
```

Vyberte jednu z možností:

- A. $O(n)$
- B. $O(\log n)$
- C. $O(n^2)$
- D. $O(1)$
- E. Jiná neuvedená možnost

Správná odpověď: A

2. Strojové učení s učitelem

2.1 Perceptron a rozhodovací hranice

Indikativní bodování: 6 bodů

Mějme 4 body v rovině rozdělené do dvou tříd:

- třída **+1**: $P_1 = (0,0)$, $P_2 = (2,2)$
- třída **-1**: $P_3 = (0,2)$, $P_4 = (2,0)$

Která z následujících možností může být rozhodovací hranicí perceptronu?

- A. $x - 1 = 0$
- B. $y - 1 = 0$
- C. $x + y - 2 = 0$
- D. $x - y = 0$
- E. Žádnou přímku nelze nalézt

Správná odpověď: E

2.2 K-nearest neighbour

Indikativní bodování: 6 bodů

Na číselné ose máme trénovací body:

- třída **A**: 0
- třída **B**: 2,3,4

Chceme klasifikovat bod 0,5 pomocí modelu K nejbližších sousedů (KNN) a jako vzdálenostní funkci použijeme absolutní vzdálenost (L1 normu, manhattanskou vzdálenost).

Uvažujeme hodnoty $k=1$ a $k=3$. Model rozhoduje prostým většinovým hlasováním.

Které tvrzení je správné?

- A. Pro $k = 1$ vyjde **A** a pro $k = 3$ také **A**.
- B. Pro $k = 1$ vyjde **A**, ale pro $k = 3$ vyjde **B**.
- C. Pro $k = 1$ vyjde **B**, ale pro $k = 3$ vyjde **A**.
- D. Pro $k = 1$ i $k = 3$ vyjde **B**.
- E. Nastane remíza pro obě hodnoty k .

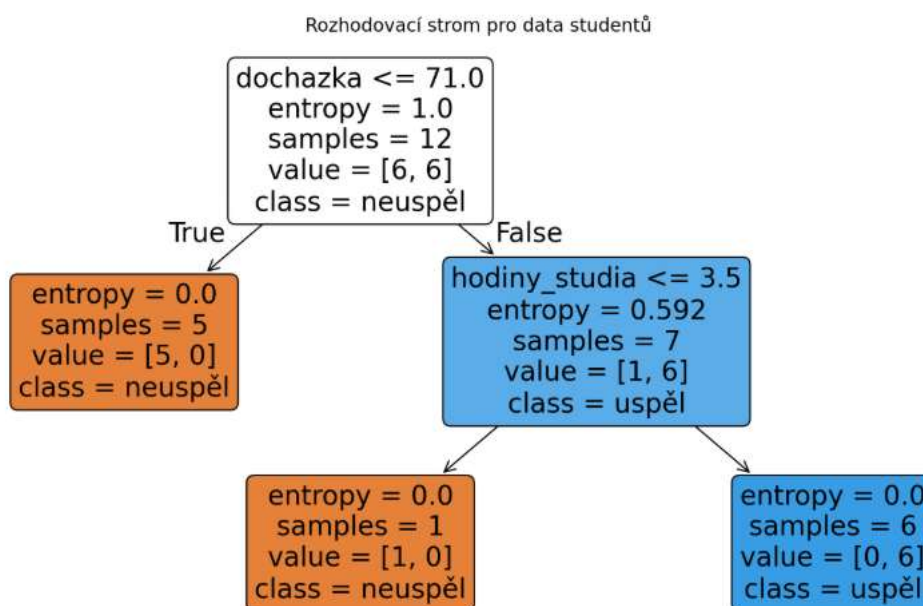
Správná odpověď: B

2.3 Rozhodovací strom

Indikativní bodování: 6 bodů

Uvažuj rozhodovací strom, který předpovídá, zda student **uspěl** nebo **neuspěl** v kurzu, na základě dvou znaků:

- **docházka** v procentech
- **počet hodin studia týdně**



Které tvrzení nejlépe vystihuje vztah v datech, který strom odhalil?

- A. O výsledku rozhoduje pouze počet hodin studia, docházka není důležitá.
- B. Nízkou docházkou lze vždy vykompenzovat vyšším počtem hodin studia.
- C. Docházka tvoří kořen stromu, protože v prvním kroku přináší nejlepší rozřazení na úspěšné a neúspěšné.
- D. Student s vysokou docházkou vždy uspěje bez ohledu na počet hodin studia.
- E. Strom ukazuje, že mezi docházkou a úspěchem není žádná souvislost.

Správná odpověď: C

3. Lineární regresní model

3.1 Klasický lineární regresní model

Indikativní bodování: 8 bodů

Předpokládejme, že na základě náhodného výběru byl metodou nejmenších čtverců odhadnut klasický lineární regresní model popisující vztah mezi cenou nemovitosti (v tisících Kč), její plochou (v m²) a typem nemovitosti (byt nebo dům). Odhadnutá regresní rovnice je:

$$\overline{cena} = 10 + 100 \text{ plocha} + 2000 \text{ dům},$$

kde *dům* je binární (dummy) proměnná, která nabývá hodnoty 1 pro dům a 0 pro byt; byt je tedy referenční (základní) kategorie.

Které z následujících tvrzení je pravdivé?

- A. Průměrná cena bytu je 110 tisíc Kč
- B. Při stejné ploše je odhadnutá cena domu o 2 miliony Kč vyšší než odhadnutá cena bytu
- C. Průměrná cena domu je 2 miliony Kč
- D. Pokud plocha nemovitosti vzroste o 100 m², tak se cena zvýší o tisíc Kč

Správná odpověď: B

4. Učení bez učitele

4.1 Asociační pravidla

Indikativní bodování: 6 bodů

V obchodě bylo analyzováno **10 nákupních košíků**. Pro pravidlo

{chleba} → {máslo}

bylo zjištěno:

- V **6** koších se vyskytl **chleba**.
- Ve **3** koších se vyskytl současně **chleba i máslo**.

Které tvrzení o hodnotách měř podpory (support) a spolehlivosti (confidence) výše uvedeného pravidla je správné a podložené daty?

- A. polovina všech zákazníků koupila chleba i máslo (podpora pravidla je 50 %).
- B. 30 % zákazníků, kteří koupili máslo, koupilo i chleba (spolehlivost pravidla je 30 %).
- C. 30 % všech košíků obsahuje chleba i máslo a 50 % košíků s chlebem obsahuje také máslo (podpora pravidla je 30 % a jeho spolehlivost je 50 %).
- D. 60 % všech košíků obsahuje chleba a 30 % z nich i máslo (podpora pravidla je 60 % a jeho spolehlivost je 30 %).
- E. 60 % všech košíků obsahuje chleba i máslo (podpora pravidla je 30 % a jeho spolehlivost je 60 %).

Správná odpověď: C

5. Evaluace modelů strojového učení

5.1 Matice záměn

Indikativní bodování: 6 bodů

Na základě výstupu binárního klasifikačního modelu byla sestavena matice záměn (confusion matrix). Co představuje hodnota **False Negative** (falešně negativní výsledky klasifikace)?

- A. Celkový počet případů, které model klasifikoval nesprávně
- B. Počet případů, kdy model nesprávně klasifikoval instance patřící do pozitivní třídy jako negativní
- C. Počet chyb modelu při klasifikaci případů náležejících do negativní třídy
- D. Podíl všech nesprávných klasifikací na celkovém počtu případů

Správná odpověď: B

5.2 Křížová validace

Indikativní bodování: 7 bodů

Které z následujících tvrzení popisuje vztah mezi **K-fold** a **leave-one-out** (LOOCV) křížovou validací?

- A. LOOCV je speciální případ K-fold validace, kdy K = počet pozorování
- B. K-fold validace je aproximace LOOCV pro malé K
- C. K-fold validace je speciální případ LOOCV pro datové sady s malým počtem pozorování
- D. Ostatní odpovědi nejsou správné, jelikož mezi LOOCV a K-fold validací neexistuje vztah

Správná odpověď: A

6. Textová analytika

6.1 Bag of Words

Indikativní bodování: 6 bodů

Uvažujme dvě věty:

- d_1 : „pes kouse clovek“
- d_2 : „clovek kouse pes“

Tyto věty převedeme do vektorové reprezentace **Bag of Words**.

Které tvrzení je správné?

- A. Vektorová reprezentace vět bude různá, protože slova jsou v jiném pořadí.
- B. Vektorová reprezentace vět bude stejná, protože Bag of Words zachycuje výskyt slov, ne jejich pořadí.
- C. Druhá věta bude mít delší vektor, protože začíná jiným slovem.
- D. Bag of Words automaticky pozná, že věty mají jiný význam.
- E. Bag of Words nelze použít na krátké věty.

Správná odpověď: B

6.2 Měření podobnosti a vzdálenosti ve vektorovém prostoru

Indikativní bodování: 6 bodů

Uvažujte dva dokumenty reprezentované vektorově jako

$$d_1 = (1, 1, 0),$$

$$d_2 = (1, 0, 1).$$

Jaká je **kosinová míra podobnosti** mezi těmito dvěma vektory?

Použijeme vzorec

$$\cos(\theta) = \frac{d_1 \cdot d_2}{\|d_1\| \|d_2\|}$$

- A. 0
- B. 0,5
- C. 1
- D. $\frac{1}{\sqrt{2}}$
- E. -0,5

Správná odpověď: B

7. Programování

7.1 Objektově orientovaný přístup

Indikativní bodování: 5 bodů

Tým vývojářů vytváří software pro malou knihovnu. V aplikaci jsou evidovány knihy, jejich výpůjčky a pak také čtenáři a zaměstnanci knihovny. Program je implementován v objektově orientovaném programovacím jazyce.

V systému existuje základní třída `User`, ze které jsou odvozeny třídy `Reader` a `Librarian`:

`User`

└─ `Reader`

└─ `Librarian`

Jaký vztah existuje mezi třídami `User`, `Reader` a `Librarian`?

- A. třídy `Reader` a `Librarian` dědí vlastnosti třídy `User`
- B. třída `User` dědí vlastnosti tříd `Reader` a `Librarian`
- C. mezi třídami existuje vztah asociace
- D. všechny třídy jsou navzájem nezávislé
- E. třídy `Reader` a `Librarian` musí být označeny jako `private`

Správná odpověď: A

7.2 Programování základní konstrukce

Indikativní bodování: 8 bodů

Tým vývojářů vytváří software pro malou knihovnu. Program potřebuje projít seznam všech výpůjček a u každé zkontrolovat, zda není po termínu vrácení. **Která programová konstrukce se pro takovou operaci typicky používá?**

- A. třída
- B. datový typ
- C. komentář
- D. cyklus
- E. import knihovny

Správná odpověď: D

7.3 Vývoj aplikace

Indikativní bodování: 5 bodů

Tým vývojářů vytváří webovou aplikaci pro správu osobních rozpočtů. Jeden z vývojářů dokončil implementaci nové funkce pro export rozpočtu do PDF ve větvi feature-export. Nyní chce tuto změnu začlenit do hlavní větve main. **Který postup je v systému Git typicky používán pro začlenění změn z jedné větve do druhé?**

- provedení operace merge mezi větvemi
- smazání větve main a následné přejmenování větve feature-export na main
- vytvoření nového repozitáře obsahujícího obě verze
- ruční kontrola obsahu souborů v obou větvích a přepsání změn ze souborů ve větvi feature-export do souborů ve větvi main
- provedení operace checkout na větvi main i feature-export

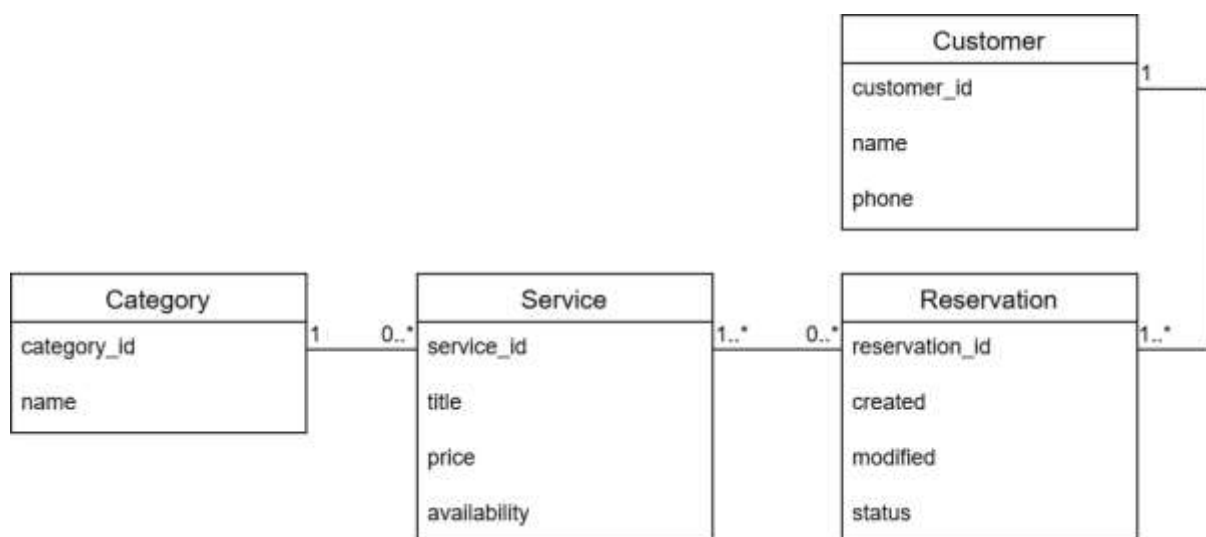
Správná odpověď: A

8. Databáze

8.1 Databáze – koncepty

Indikativní bodování: 5 bodů

Uvažujme konceptuální datový model systému pro správu online rezervací služeb. Systém eviduje zákazníky, rezervace, nabízené služby a jejich rozdělení do kategorií.



K dispozici máme také textový popis vazeb:

- kategorie může obsahovat libovolné množství služeb
- služba je vždy v jedné kategorii
- zákazník má jednu či více rezervací

- každá rezervace je přiřazena právě jednomu zákazníkovi
- rezervace obsahuje jednu či více služeb
- služba může být zařazena do více rezervací, ale nemusí být v žádné

Na obrázku vidíte konceptuální model.

Jaký je typický způsob implementace vztahu M:N mezi Reservation a Service v běžné relační databázi (např. Oracle, MariaDB, MySQL atp.)?

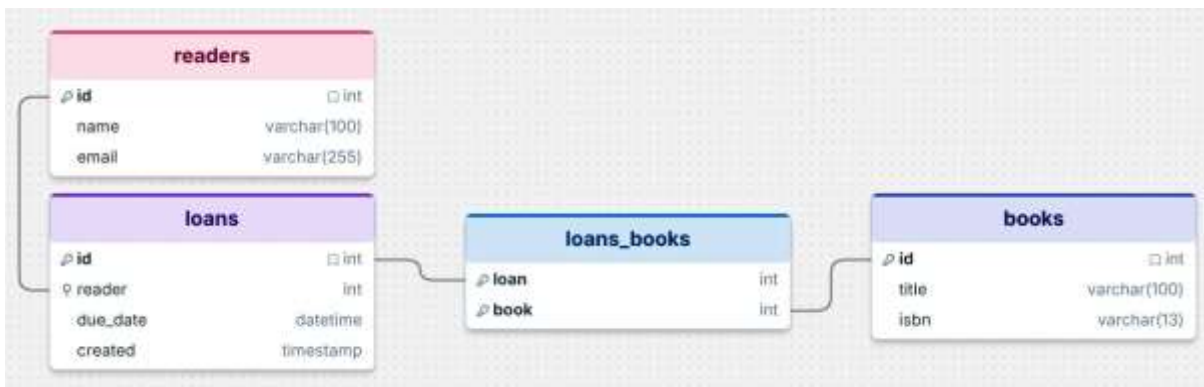
- přidáním service_id do tabulky Reservation
- sloučením tabulek Reservation a Service do jedné
- uložením seznamu služeb jako textu do tabulky Reservation
- přidáním reservation_id do tabulky Service
- vytvořením vazební tabulky odkazující na Reservation a Service

Správná odpověď: E

8.2 Structured Query Language (SQL)

Indikativní bodování: 6 bodů

Podívejte se na přiložené schéma fyzického modelu relační databáze jednoduché knihovny. V žádném ze sloupců nejsou povolené null hodnoty.



Který SQL příkaz vrátí jména a id čtenářů a počet jejich výpůjček?

- SELECT readers.id, readers.name, COUNT(loans.id) FROM readers LEFT JOIN loans ON readers.id = loans.reader GROUP BY readers.name, readers.id;
- SELECT readers.name, COUNT(books.id) FROM books JOIN loans ON books.id=loans.id GROUP BY readers.name, readers.id;
- SELECT loans.reader FROM loans JOIN readers ON loans.reader = readers.id;
- SELECT COUNT(readers.id, readers.name), COUNT (loans.id) FROM readers LEFT JOIN loans ON readers.id=loans.reader;

E. SELECT readers.id, readers.name, COUNT (loans_books.book) FROM readers LEFT JOIN loans ON readers.id=loans.reader LEFT JOIN loans_books ON loans_books.loan=loans.loan GROUP BY readers.name, readers.id;

Správná odpověď: A