

Baigiamasis Data Science atsiskaitymas

Dalykas: Data Science [CADS2]

Atsiskaitymo data: Kovo 27-28 dienos (8 akademinės valandos)

Atsiskaitymo forma: testas ir baigiamasis projektas

Vertinimo skalė: 1-10 [testas] ir 1-10 [baigiamasis projektas]. Kiekvienos dalies pirmasis teigiamas pažymys yra 7. Atsiskaitymas laikomas sėkmingu tik jei iš abiejų dalių yra surenkamas teigiamas balas.

Baigiamasis teorijos atsiskaitymas

Kovo mėn. 27 d. vyks teorijos atsiskaitymas. Pradžia: kovo mėn. 27 d. 18.10 val. Pabaiga: kovo mėn. 27 d. 19.25 val. Trukmė: 75 minutės. Atsiskaitymo forma: testas su 20 klausimų [kiekvienas vertas po 1 tašką] + 6 papildomi klausimai [verti po 0.5 taško]. Maksimalus taškų skaičius: 23 taškai. Testas laikomas išlaikytu surinkus 14 ir daugiau taškų.

Galimi įvairaus pobūdžio klausimai:

- su duotais 4 galimais atsakymais [galimi keli teisingi atsakymo variantai]
- tiesa/netiesa tipo klausimai
- atviro tipo testo klausimai

Klausimai sukurti pagal kurso temas (Programavimas su Python, Tiriamoji duomenų analizė, Mašininis mokymasis, Duomenų bazės), testo metu galima naudotis visais šaltiniais.

Pavyzdiniai testo klausimai

1. Mašininis mokymasis:

Kas yra ansamblio mokymasis (angl. *ensemble learning*)? Kokius žinote šio mokymosi tipus? Trumpai paaiškinkite kiekvieną iš jų.

2. Mašininis mokymasis:

Persimokymas (angl. *overfitting*) - viena iš pagrindinių problemų mokant mašininio mokymosi modelį. Modelis tampa pernelyg geras mokymosi stadijoje, tačiau naujiems duomenis modelio tikslumas yra mažas. Kokie galimi persimokymo problemos sprendimo būdai:

- a) regularizacija (angl. *regularization*);
- b) sukurti paprastesnį ir mažiau parametrų turintį modelį;
- c) naudoti giliojo mokymosi (angl. *deep learning*) algoritmus;
- d) duomenų augmentavimas (angl. *data augmentation*).

3. Python

Koks bus žemiau pateikto kodo rezultatas?

```
def outer_fun(a, b):  
    def inner_fun(c, d):  
        return c + d  
    return inner_fun(a, b)  
  
res = outer_fun(5, 10)  
print(res)
```

- a) (5, 10);
- b) 15;
- c) Syntax Error;
- d) Error calling function.

4. EDA

Koks rezultatas bus atspausdintas įvykdžius kodo bloką?

```
import pandas as pd

l = [
    ['a', 12, 12],
    [None, 12.3, 33.],
    ['b', 12.3, 123],
    ['a', 1, 1]
]

df = pd.DataFrame(l, columns=['a', 'b', 'c'])

print(df.groupby(by='a', dropna=False).sum().sum().sum())
```

- a) 206.6;
- b) 161.3;
- c) Error calling sum function;
- d)

b	37.6
c	169.0

5. Duomenų bazės

Kas yra Common table expression?

- a) Laikina lentelė turinti vardą ir egzistuojanti pagrindinės užklausos vykdymo metu.
- b) Sąlyga padedanti filtruoti norimas eilutes.
- c) Group by būdas, kuris iš karto išrikiuoja rezultatą.

Baigiamasis projektinių darbų atsiskaitymas

Kovo mėn. 27-28 dienomis vyks projektų rezultatų pristatymas kur kiekvienas studentas pristatys pasirinktą darbo metodiką, projektą ir rezultatus. Pradžia: kovo mėn. 27 d. 20.00 val. Kiekvieno studento projekto aptarimui skiriamos 35 minutės:

- 15-20 minučių projekto pristatymui;
- 15-20 minučių Q&A daliai.

Norint gauti teigiamą balą reikia surinkti 7 taškus. Maksimalus taškų skaičius yra 12 taškų:

- 4 taškai skiriami už sėkmingai įgyvendintą projektą;
- 3 taškai skiriami už projekto pristatymą [įskaitant Q&A];
- 1.5 taškas skiriamas už kodo repozitorijos struktūrą [README.md, aplinkos failas];

- 1.5 taškas skiriamas už kodo tvarkingumą pagal gerąsias PEP8 praktikas; Galima naudoti black formatavimo paketą.
- BONUS: 2 taškai už aktyvų dalyvavimą kitų pristatymuose.

Q&A forma:

Prieš prasidedant projekto pristatymui, atsiskaitomajam bus priskirti du studentai kurių tikslas pasibaigus projekto pristatymui užduoti mažiausiai po du klausimus apie pranešimo turinį ar pasirinkimus. Už aktyvų dalyvavimą klausiant - skiriami bonus taškai [žiūrėti vertinimo skalę]. Tokio atsiskaitymo tikslas - kiekvienas studentas turės galimybę nors keletą kartų pabūti *stakeholderiu* ir susipažintų ne tik su savo, tačiau ir kitų studentų sprendimais, gebėtų pritaikyti išmoktas žinias dideliame spektrui ML problemų.