

Baigiamojo darbo reikalavimai

Projektų pridavimo data: gegužės 13 diena.

Projektai turi būti GitHub repository su aišku aprašymu README.md. Repository gali būt ir privati ir vieša pagal poreikį.

Repository turi būti:

- Readme.md
 - Projekto aprašymas (jei darote asmeninį projektą, jei darote duotą, užtenka projekto pavadinimo)
 - Modelio pasirinkimas su galutinėmis metrikomis.
 - Instrukcijos kaip pasileisti API serverį.
 - Jei dirbate su savais duomenimis, prašau pridėkite testavimui skirtus duomenų pavyzdžius. Kas dirba su konfidencialiais duomenimis, prašau pateikite dirbtinius duomenis kurie atitiktų realių duomenų formatą.
 - Pasiektas rezultatas, išvados.
- Notebook su duomenų analize ir modelio treniravimo, spėjimo (prediction) kodu.
 - Paleidus notebook – jis įsivykdo be klaidų nuo pradžios iki pabaigos
 - Pagrindiniai žingsniai turi būti atskirti ir paaiškinti.
 - Rezultatai atskirti ir paaiškinti.
- Model deployment kodas su išsaugotu supakuotu (pickle) modelio failu.
- Kodas kurį naudojote scrapinti duomenis (jei dirbote su scrapinimu)

Vertinimo kriterijai:

1. Duomenų paruošimas – 20%:

- 0. Duomenų analizė ir jos pilnumas
 - 0. Duomenų tvarkymas - „missing values“, outliers, kategoriniai kintamieji ir t.t.

2. Duomenų analizės 20%:

- 0. aprašymas ir taikomų prielaidų tinkamumas/pagrįstumas
 - 0. Stulpelių statistiniai duomenys
 - 1. Histogramos
 - 2. Kokias galite padaryti prieladas ar išvadas

3. Modelis 20%:

- 0. Modelio pasirinkimas ir tinkamumas
- 1. Modelio pajėgumas („performance“) – kuo didesni modelio metrikai, tuo geriau. Dirbantiems su klasteringu („clustering“) – klasterių interpretacija pakeičia pajėgumą kaip vertinimo kriterijų.
 - 0. Turėti benchmark modelį.
 - 1. Modelio pajėgumas yra vertinamas atsižvelgiant, kiek jis yra geresnis už benchmark modelį.

4. Modelio „deployment“ 20%:

- 0. API veiksnumas
 - 0. Išsaugom modelį kaip binarį failą (pkl formatas)
 - 1. Sukuriam Docker image su mūsų modeliu
 - 2. Paleidžiam docker image ir galime gauti modelio prognozes, jį kviesdami per API

5. Kodo kokybė 20%:

- 0. Kodo aiškumas, švarumas
 - 0. Paparasti ir pasikartojantys kodo gabaliukai sudėti į funkcijas
 - 1. Kodas nepasikartojantis – DRY – don't repeat yourself.

6. Prezентacija 20%:

- 0. Žodinė – aiškumas, konkretumas
- 1. Raštinė – aiškumas, pilnumas

7. Bonus 10%:

- 0. Neaprašytas papildomas darbas