

Velika domača naloga - Bicikelj

UOZP

Jan Zorko, 63190324

29. maj 2023

1 Opomba

Zaradi omejitev nalaganja datotek nisem mogel oddati vseh potrebnih datotek, tako da oddan projekt ni celoten. Oddajo sem prilagodil tako, da se požene brez njih, vendar tako seveda ne proizvede enakih rezultatov. Predvsem manjkajo podatki o vremenu in razdaljah.

Celoten projekt je zato na voljo kot GitHub repozitorij: <https://github.com/JZorko-FRI/uozp-bicikelj>

2 Uporaba

Vstopna točka je datoteka `main.py`, datoteka `readme.md` pa na hitro opiše še ostale, saj je koda razdeljena po delih. Če podamo parameter `predict` se napovedi testne množice zapišejo v `predictions.txt`.

Alternativno lahko podamo parameter `evaluate`, kar izpiše MAE za vsak posamezen model na ročno določeni testni množici. Ta vsebuje podatke v urah 8, 20, in 21, medtem ko učna množica vsebuje vse ostale.

3 Značilke

Poleg predlaganih značilk so dodane še:

- Vreme (da/ne podatek za dež, nevihta, sneg, poledica)
- Trajanje poti v sekundah do 10 najbližjih postaj (Google Maps API, `analysis/maps.py`)
- Ročno določen faktor prometa, visok v času prometnih konic
- Število koles na postaji pred enim tednom in po enem tednu
- Binariziran mesec in binariziran dan v tednu
- Binarizirana kombinacija 15 min. intervala v dnevu in dneva v tednu
- In še nekaj drugih, manj vplivnih, vidnih v kodi...

4 Model

Končna rešitev uporablja XGBoost, s parametri: `n_estimators=50`, `eta=0.1`, `max_depth=5`, `subsample=0.8` in `colsample_bytree=0.8`.

Ta doseže rezultat MAE na prvi polovici 1.92473.

Grajeni so ločeni modeli za vsako postajo. Dodatne ločitve po mesecu ali dnevu v tednu so bile slabše.