

Velika domača naloga UOZP: Bicikelj

Ema Leila Grošelj

February 16, 2025

1 Uvod

Na Bicikelj postajah se pogosto zgodi, da na kakšni postaji koles sploh ni - ali pa obratno, da je postaja polna in kolesa več ne gre zakleniti. Upravitelj Bicikelja se proti temu bori tako, da s kombiji kolesa razvaža med postajami. Moj cilj je bil narediti model, ki bi upravitelju pomagal oceniti, katera kolesa mora kam prestaviti, tako da pravilno napovem stanje na postajah, čez eno in dve uri. No voljo sem imela približno za dva meseca podatkov (lanski avgust in september) - število koles na vseh Bicikelj postajah za vsake 5 minut.

2 Metode

Nalogo sem izdelala v Pythonu. Za model sem izbrala `gradientBoostingRegressor` iz knjižnice `sklearn` in zaokrožila njegove napovedi na najbližje število. Za delo s podatki sem uporabila še `numpy` in pa `pandas`. Podatke o vremenu sem pridobila s knjižnico `meteostat`.

3 Rezultati

3.1 Model

Že test na prvi postaji s prečnim preverjanjem me je prepričal, da bo GBR boljša izbira za model, saj je imel dvakrat manjšo napako kot linearna regresija.

3.2 Značilke

Moje značilke so bile naslednje:

- `jePonedeljek[bool]`
- `jePetek[bool]`
- `jeVikend[bool]`
- `pocitnice[bool]`
- `minutaDneva[int]`
- `stKolesPredEnoUro[int]`

- `stKolesPredDvemaUrama[int]`

Za binarne značilke dni v tednu npr. `jePonedeljek` sem se odločila, ker sem sklepala, da sta ponedeljek in petek posebna dneva, pravtako vikend. Za minuto v dnevu sem se odločila, zato da obdržim čim večjo zveznost časa in lahko model sam presodi, kdaj se recimo frekvenca uporabe poveča.

Število koles sem nastavila na časovno najbližji še znani podatek, tako sem zapolnila neznane podatke. Potem sem opravila test z prečnim preverjanjem na testnih podatkih. Zanimalo me je koliko postajam je dodan atribut `stKolesPredUroInPol` zmanjšal napako. Dobila sem da se je na 48 odstotkov postaj napaka povečala na 52 odstotkih pa zmanjšala. Zato ker ni prinesel opaznega izboljšanja in ker je pri polovici testnih primerov resničen podatek neznan, sem se odločila da atributa ne uporabim. Po drugi strani se je to, da sem poleg atributa `stKolesPredEnoUro` dodala še `stKolesPredDvemaUrama`, izkazalo za dobro, saj se je povprečna napaka prečnega preverjanja zmanjšala na prav vseh postajah.

Prav tako nisem uporabila vremena. Če sem vreme gledala za 1 uro nazaj (kar bi prav prišlo tudi v praksi, da se izognemo napaki vremenske napovedi) je bil rezultat boljši kot brez zamika. Prav tako je bil boljši, če sem namesto vseh podatkov o vremenu uporabila je podatka o dežju in temperaturi. A izkazalo se je, da je rezultat, če vreme zanemarim, še najboljši. Delno bi to pripisala premajhnemu deležu res deževnih dni v podatkih in potencialni neodvisnosti uporabe koles od temperature.

4 Zaključek

Moja najboljša končna točnost na vidnem delu testov, je bila 1.82169 kolesa. Nad točnostjo sem presenečena. Zdi se mi da bo model dostavljalcem koles, v naslednjem avgustu in septembru pomagal, v kolikor bo podobno vreme. Menim pa tudi, da ne bi škodilo, če bi raje napovedovala spremembo (zahteve uporabnikov), saj bodo tudi premeščevalci koles vplivali na število koles na postajah.

