

# Kalibracija meritev pametnih svetilk s strojnim učenjem

Marko Hostnik

## Uvod

Za projekt pri predmetu Matematika z računalnikom rešujem problem kalibracije meritev pametnih svetilk. Pri tem je cilj uporabiti metode nadzorovanega in nenadzorovanega strojnega učenja. Pobudo za projekt je dalo podjetje Garex<sup>1</sup>, ki v sodelovanju z drugimi podjetji razvijajo rešitev za *pametno* mestno razsvetljavo. S projektom sem se vključil tudi na tekmovanje, ki ga organizira Fakulteta za računalništvo in informatiko<sup>2</sup>. V ekipi z mano sodeluje Gregor Kikelj, študent 3. letnika dodiplomskega študija Matematike na FMF.

## Opis problema

Pametne svetilke za mestno razsvetljavo niso le svetilke, temveč so opremljene z različnimi merilnimi napravami. Vsaka svetilka vsebuje merilnike temperature, tlaka, vlage, onesnaževanja zraka (CO<sub>2</sub>, PPM) ter radar za detekcijo pešcev in avtomobilov. Vendar vse meritve niso zanesljive, zato jih želimo kalibrirati. Ker so svetilke geografsko blizu (na isti ulici) si lahko pomagamo z združenimi podatki vseh svetilk. Trenutno imamo na voljo tri svetilke, ki se približno eno leto nahajajo v Mariboru. Cilj je uporabiti zbrane podatke in jih obdelati z metodami strojnega učenja v namen približati se k resničnim vrednostim izmerjenih količin.

## Načrt dela

Za delo uporabljamo programski jezik Python in programske knjižnice za strojno učenje ter analizo podatkov (*scikit-learn*, *matplotlib*, *pandas*, ...).

### Delo do sedaj

Izvedli smo uvodno analizo podatkov ter pripravili ogrodje za testiranje naučenih modelov strojnega učenja.

### Obdelava podatkov

Najprej smo dobili dostop do podatkov svetilk in jih obdelali v obliko primerno za obdelavo. Ker ima vsaka meritev 68 spremenljivk, smo se odločili, da se osredotočimo najprej na vremenske podatke. Za primerjavo z resničnimi vremenskimi podatki smo si pomagali z arhivskimi vremenskimi podatki ARSO<sup>3</sup> iz samodejne postaje *Maribor – Urbanski plato*.

Ugotovili smo, da meritve temperature (in drugih količin) niso stabilne. Včasih se izmerjena temperatura povzpne tudi do 600 °C. Zato smo očitno nesmiselne podatke odstranili

iz zbirke meritev. Omenimo tudi, da se meritve izvajajo na približno 30 minut.

Za uvodno analizo smo združili podatke vseh svetilk in analizirali podatke za vsak mesec posebej. Ugotovili smo, da svetilke poročajo višjo temperaturo od resnične. Ta sprememba je bolj izrazita v zimskih mesecih kot v poletnih. Ker bi bila možna razlaga za odstopanje temperature v izpostavljenosti soncu, smo podatke združili tudi po uri (24 ur) in ugotovili, da je temperatura tudi ponoči previsoka. Primer je viden na Sliki 1.

Podobno smo analizirali tudi vlago in tlak, kar sedaj izpuštimo iz poročila.

### Strojno učenje

Ključni del modeliranja podatkov je preverjanje kako se bo model obnesel v praksi. Zato smo implementirali programsko ogrodje za testiranje napovedi modelov. Pri tem je pomembna skrb za pravilno metodologijo testiranja, kajti podatki so časovne narave. To pomeni, da ko testiramo model na nekem časovnem odseku, moramo model učiti samo na podatkih, ki so kronološko pred časom vseh meritev v testnem odseku. Rezultate napovedi modelov ovrednotimo z izračunom povprečnega kvadratnega odstopanja (MSE) in povprečnega absolutnega odstopanja (MAE) od resničnih vrednosti.

Za primerjavo različnih modelov smo implementirali nekaj osnovnih različic, kot so napovedovanje povprečja, mediane, ali pa odštevanje povprečne napake.

### Nadaljnje delo

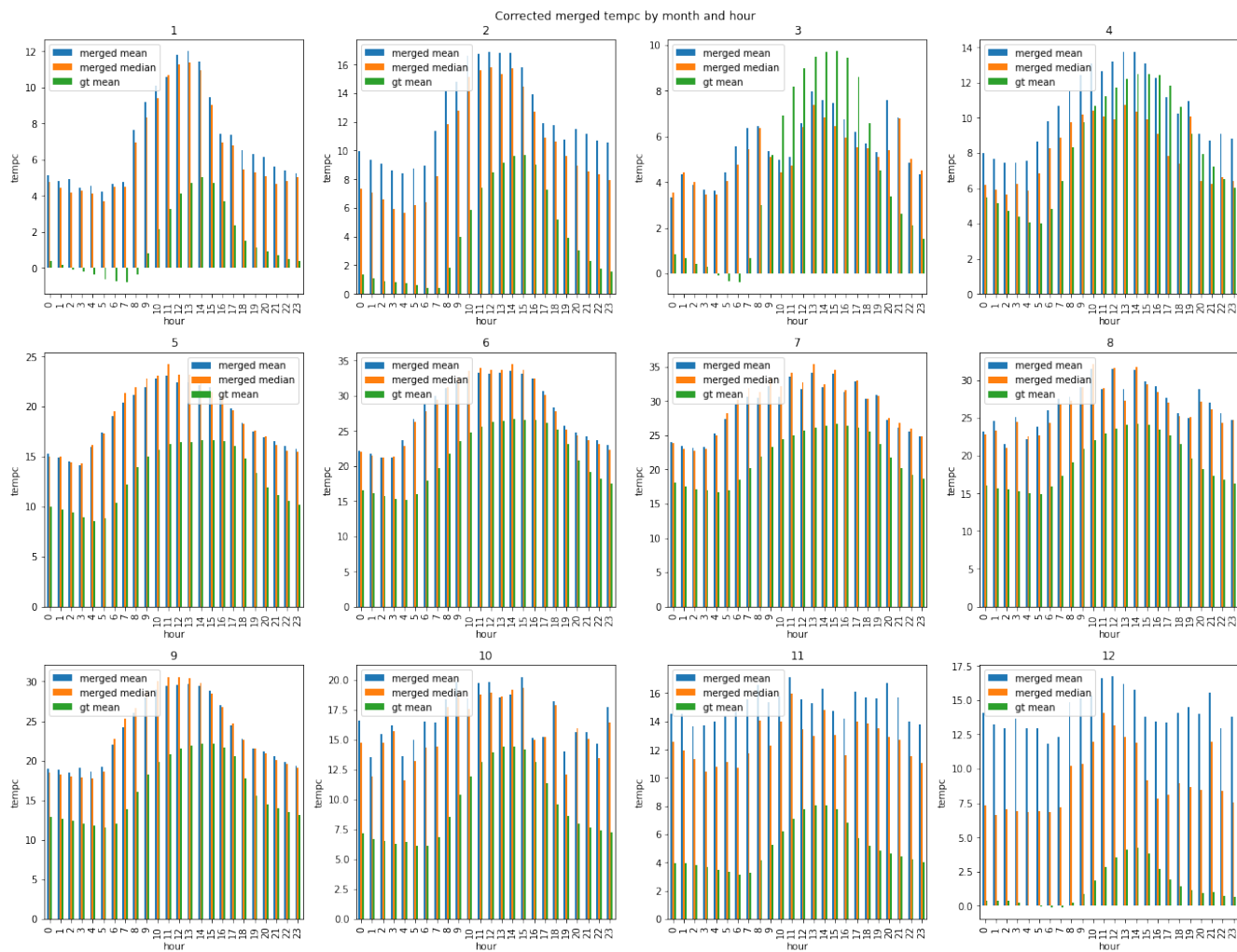
V nadaljevanju bomo preizkusili različne modele strojnega učenja. Začeli bomo z bolj preprostimi in šli proti kompleksnejšim. Ključna dva modela bosta linearna regresija in na ključni gozdovi, saj sta zelo robustna. Pri tem bo pomembna izbira neodvisnih spremenljivk meritev na podlagi katerih bomo učili model, ki bo napovedal ciljno spremenljivko. Na primer, zagotovo bo pri napovedi temperature koristila informacija o času meritve. Posledično bo potrebno opraviti mnogo poskusov z različnimi kombinacijami vhodnih spremenljivk in načinom predobdelave njihovih vrednosti.

Do sedaj smo z modeli napovedovali samo temperaturo, v prihodnje bomo pa preizkusili tudi ostale količine, kot sta tlak in vlaga.

<sup>1</sup><https://www.garex.si/>

<sup>2</sup><https://datascience.fri.uni-lj.si/competition/>

<sup>3</sup><https://meteo.arso.gov.si/>



**Slika 1.** Primerjava resnične in izmerjene temperature po mesecih. Zelena barva (*gt*) označuje resnične vrednosti, modra in oranžna pa povprečje in mediano združenih meritev od vseh naprav.