

Kalibracija meritev pametnih svetilk s strojnim učenjem

Matematika z računalnikom

Marko Hostnik

Maj 2022

Uvod

- FRI DataScience tekmovanje
- podjetje Garex
- *pametna* mestna razsvetljava



Opis problema

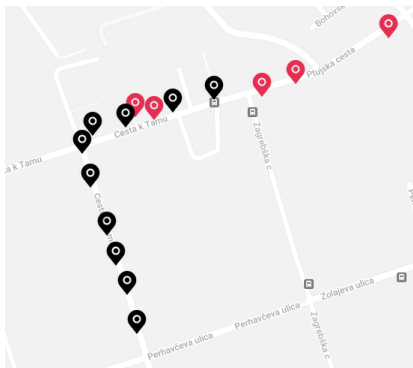
- senzorji za **temperaturo**, tlak, vlago, onesnaževanja zraka (CO₂, PPM), radar za detekcijo pešcev in avtomobilov
- nezanesljive meritve \Rightarrow **kalibracija**
- marec 2021 – marec 2022 \Rightarrow **3** svetilke
- marec 2022 – zdaj \Rightarrow **10** svetilk

Cilj

- analiza podatkov
- kalibracija s strojnim učenjem

Analiza podatkov

- Python (*scikit-learn, pandas, matplotlib ...*)
- vsaka meritev **68** spremenljivk \Rightarrow samo vreme
- ARSO: *Maribor – Vrbanski plato*

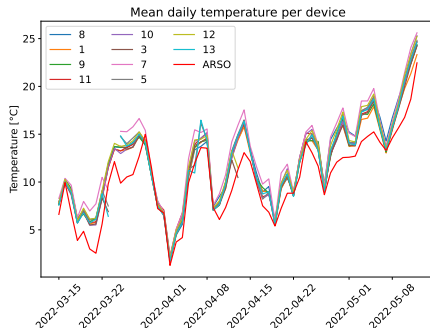
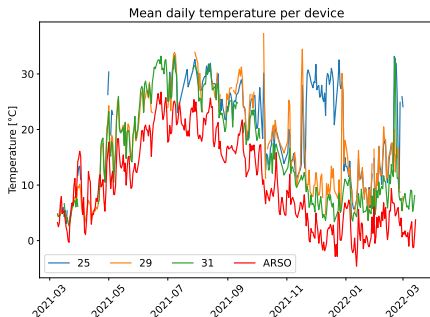


Ugotovitev

Novi podatki so bolj zanesljivi od starih!

Temperatura

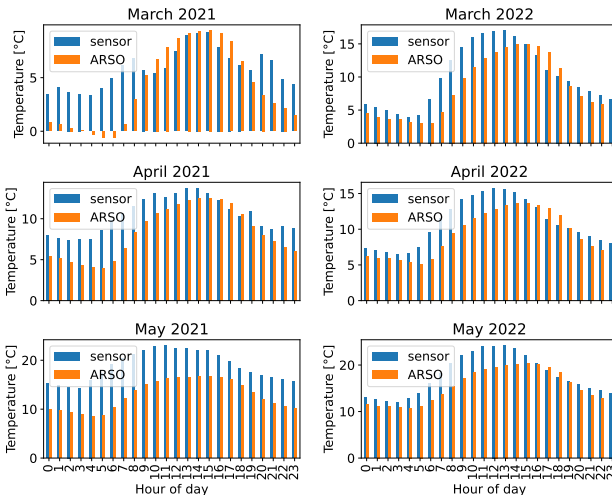
- Stari podatki: anomalije (do 600 °C)
- ARSO temperatura je nižja



Slika: Dnevna povprečna temperatura.

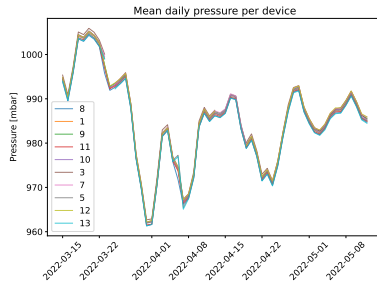
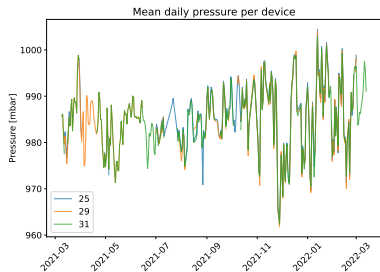
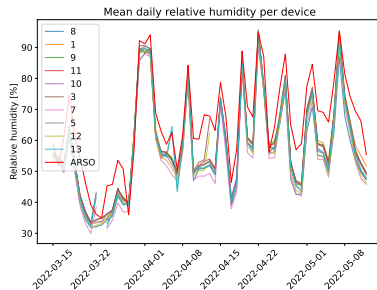
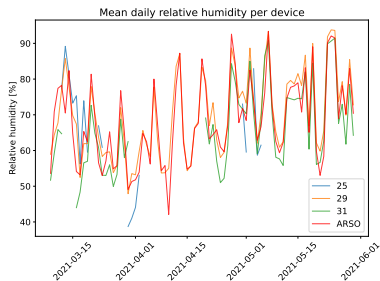
Temperatura

- Izpostavljenost soncu?



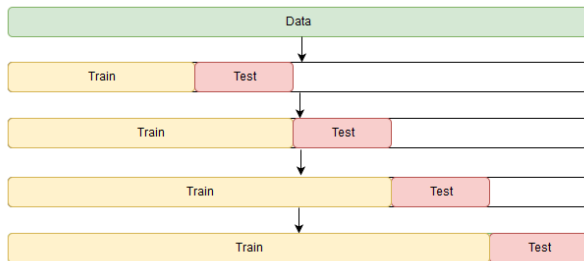
Slika: Povprečna temperatura po urah v mesecih. Levo stare, desno nove.

Vlaga in zračni tlak



Modeliranje s strojnim učenjem

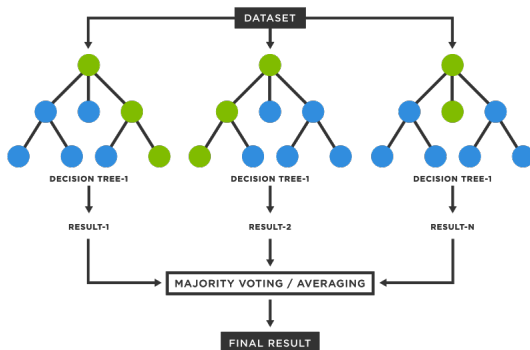
- Ogrodje za testiranje
- Časovno prečno preverjanje



- $MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2$
- $MedAE = median(|\hat{y}_1 - y_1|, \dots, |\hat{y}_n - y_n|)$
- Standardna napaka z *bootstrapping*

Modeli

- *Baselines*: identiteta, povprečje, mediana, drseče povprečje, povprečno odstopanje ...
- **Ridge** = linearna regresija z L2 regularizacijo
 - $\hat{y}(x) = \beta_0 + \sum_{i=1}^n \beta_i x_i = \beta^T x$
 - minimiziramo $J(\beta) = \sum_{i=1}^m (\hat{y}_i - y_i)^2 + \lambda \sum_{i=1}^n \beta_i^2$
- **Naključni gozdovi**



Eksperimentalna evalvacija

- hiper-parametri, pred-procesiranje podatkov (standardizacija, povprečenje ...)
- Čas kot krožnica (\sin, \cos) ali *one-hot encoding*

Rezultati

	MSE	MedAE	< 1 °C	< 3 °C	< 5 °C
Naključni gozdovi	14.144 ± 0.097	2.210 ± 0.011	0.245	0.621	0.828
Ridge	21.916 ± 0.157	2.624 ± 0.014	0.212	0.553	0.760
Drseče povp. odst.	31.586 ± 0.219	3.452 ± 0.017	0.155	0.439	0.658
Povp. odst.	50.713 ± 0.464	3.097 ± 0.019	0.179	0.488	0.682
Povprečje	80.277 ± 0.315	7.398 ± 0.026	0.065	0.201	0.333
Mediana	82.048 ± 0.328	7.400 ± 0.042	0.056	0.188	0.315
Drseče povp.	83.291 ± 0.405	7.034 ± 0.017	0.014	0.082	0.244
Identiteta	102.464 ± 0.755	6.330 ± 0.016	0.022	0.106	0.319

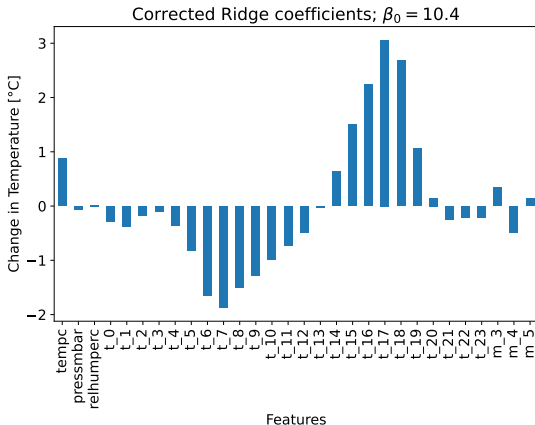
Tabela: Rezultati za napoved temperature [°C] na **starih** podatkih.

	MSE	MedAE	< 1 °C	< 3 °C	< 5 °C
Ridge	4.163 ± 0.016	1.214 ± 0.003	0.414	0.877	0.974
Drseče povp. odst.	5.938 ± 0.015	1.697 ± 0.003	0.268	0.778	0.964
Povp. odst.	7.891 ± 0.027	1.797 ± 0.004	0.262	0.745	0.922
Naključni gozdovi	8.547 ± 0.031	1.672 ± 0.004	0.322	0.737	0.910
Drseče povp.	9.262 ± 0.028	1.853 ± 0.007	0.326	0.675	0.874
Identiteta	11.229 ± 0.040	1.900 ± 0.007	0.318	0.680	0.861
Povprečje	39.532 ± 0.117	4.011 ± 0.009	0.135	0.381	0.575
Mediana	40.420 ± 0.118	4.100 ± 0.022	0.127	0.374	0.574

Tabela: Rezultati za napoved temperature [°C] na **novih** podatkih.

Interpretacija

- Prostor za izboljšave ...
- Novi podatki bolj zanesljivi \implies preprostejši modeli zadoščajo
- Ridge je enostaven, robusten, interpretabilen model!



Vprašanja?

<https://github.com/mh9533/matrac-2022>