

Bicikelj poročilo

Aleks Stepančič

Maj 2023

1 Priprava značilk

Za vsak primerek (t_i, p_j) ob času t_i na postaji p_j sem napovedovali dva tipa napovedi, čez eno uro $N1$ in čez dve uri $N2$. Smiselno sem zgradil naslednje značilke:

1. (a) Če gre za napoved tipa $N1$, smo dodali vrednost na postaji p_j pred eno in dvema urama
(b) Če gre za napoved tipa $N2$, smo dodali vrednost na postaji p_j pred dvema in tremi urama

Te vrednosti poiščemo, da od podanega časa primerka odštejemo eno, dve ali tri ure in najdemo primerke z najbližjim časom, čigar vrednost postaje p_j vnesemo. To prinese tudi primere, ko je vnešena vrednost pred eno uro enaka trenutni vrednosti primerka, saj je ta najbližja.

2. Binarno vrednost ali smo meseca avgusta ali ne, saj avgusta so šolske počitnice, kar vpliva na uporabo koles.
3. Iz <https://meteo.arso.gov.si/> sem prejel podatke o padavinah v ustreznem obdobju. Vsak čas primerka sem preslikal v najbližji čas meritve padavin. Če je vrednost padavin manj kot 0.5mm, sem vrednost nastavil na nič, drugače pa na ena.
4. Štiri $f(x) = \sin(ax + b)^2$ za $b \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$, kjer je x sekunda v dnevu in a tak, da ima $f(x)$ periodo enako enemu dnevu.
5. Dan v tednu - one-hot-encoded
6. Ura - one-hot-encoded

2 Model

Za vsako postajo in vsak tip napovedi $N1$ in $N2$ sem zdravil svoj model. Preizkusil sem Ridge regresijo ter GradientBoostingRegression s katerim sem tudi dosegel najboljši rezultat. Po predikcijah modela sem vrednosti, ki so bile negativne postavil na 0 in vrednosti, ki so presegale maksimume postavil na maksimum postaje. Napovedi sem zaokrožil na najbližje celo število.