

# 1 Opis modela

## 1.1 Arhitektura

Za vsako postajo je bil izdelan in naučen svoj model. Vhodni podatki so vsebovali naslednje značilke:

- **24 zakasnitev** (station\_t-1 do station\_t-24) trenutne postaje,
- **1 zakasnitev sosednjih postaj** ( $k = 2$ ),
- **3 vremenske značilke**: temperature, precipitation, wind\_speed,
- **8 časovnih značilk**: hour\_sin, hour\_cos, dow\_sin, dow\_cos, month\_sin, month\_cos, is\_weekend, is\_holiday.

Značilke, ki so predstavljale število koles in vremenske značilke so bile normalizirane glede na učno množico.

Arhitektura vseh modelov je bila enaka:

```
nn.Linear(37 → 128)
ReLU
nn.Linear(128 → 64)
ReLU
nn.Linear(64 → 1)
```

## 2 Vrednotenje modelov

Prvotni podatki so bili razdeljeni v razmerju 8:1:1 na train, validation in test množice. Validation in test množici posnemata strukturo dejanskih testnih podatkov, okna velikosti 48+4. Modeli so se med treniranjem vrednotili glede na napako na validation množici. Končno, lokalno, testiranje modelov je bilo opravljeno na test množici, celotna povprečna izguba modelov je bila 14.41304. Napaka se je med modeli precej razlikovala, od 2.17311 do 64.5397.

## 3 Razlaga modela

Ker ima vsaka postaja svoj model, je v predstavitvi vključen le eden izmed SHAP diagramov, vendar so ugotovitve splošne za vse postaje.

Glavni prispevki:

- Največji vpliv ima **t-1** vrednost ciljne postaje.
- Število **t-1 najbližjih sosedov** ( $k=2$ ).
- **Ura v dnevu** (hour\_sin/hour\_cos).
- Zgodovinski podatki ciljne postaje
- Nekatere vremenske značilke.

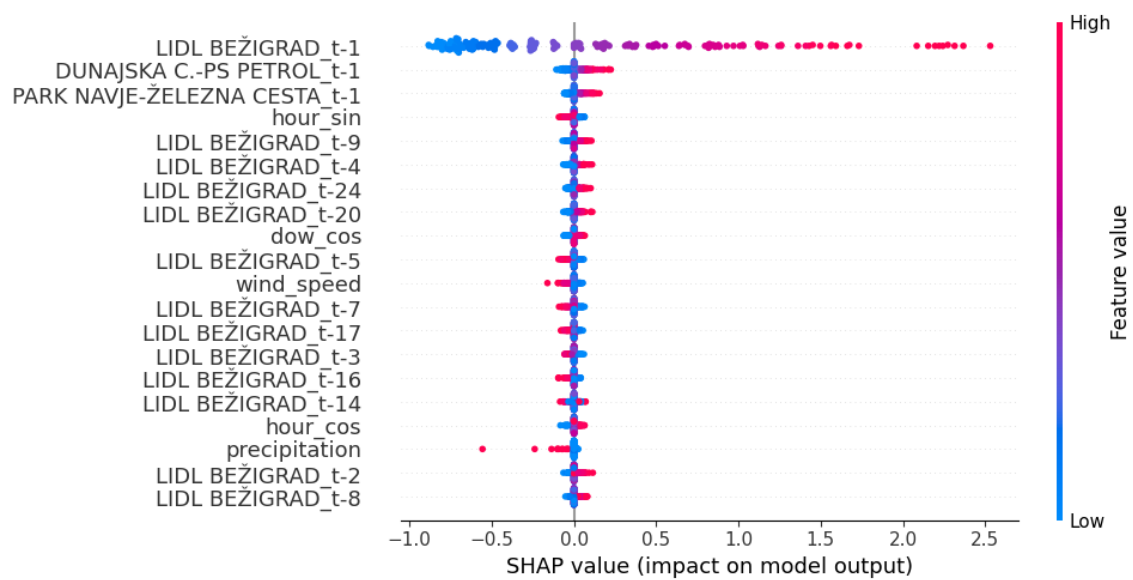


Figure 1: SHAP diagram 1. postaje