

1 Tekmovanje

1.1 Atributi

Atributi, ki sem jih uporabil pri tekmovanju so naslednji:

- Dnevi tedna - Uporabljen samo 1 stolpec, vrednosti so v intervalu 1-7 (ponedeljek = 1, torek = 2,...)
- Čas - Uporabljeni 3 stolpci. Prvi predstavlja uro in minuto odhoda združeno skupaj (16h 30m -> 16.5), drugi stolpec je kombinacija ure in dneva odhoda (ura * dan * 6), pomnoženo je s 6, da dobim bolj podroben interval. Vrednosti so na intervalu 2-1008. Zadnji, tretji stolpec pa predstavlja uro odhoda (interval 1-24).
- Počitnice in dela prosti dnevi - 1 stolpec za počitnice, 1 stolpec za državne praznike (vrednosti 1 ali 0). Skupaj 2 stolpca.
- Dež - Če je povprečna dnevna količina padavin presegla mejo 33 mm, je v stolpcu 1, sicer 0. Skupaj 1 stolpec.
- Vozniki - Zakodirani z one-hot kodiranjem. Prisotnost voznika za določeno vožnjo je označena z 1 v ustreznem stolpcu. Skupaj 430 stolpcev.

Skupno sem uporabil 437 atributov. K dobljenem rezultatu so največ prispevali atributi dnevi v tednu, ter časovni atributi. Treniral pa sem samo na podatkih voženj iz novembra.

1.2 Implementacija in napovedni modeli

Za nalogo sem uporabil 2 modela, to sta XGBRegressor in XGBRFRegressor iz knjižnice [xgboost](#). Tako kot pri prejšnji nalogi, sem naredil za vsako linijo svoj model. Za iskanje najboljšega modela za trenutno linijo sem uporabil 10-kratno prečno preverjanje. Poskušal sem tudi z ostalimi modeli (odločitvena drevesa, nevronske mreže..), vendar sta mi zgoraj omenjena modela prinesla najboljši rezultat glede na moje attribute.

K boljši napovedi pa je pripomoglo tudi zlaganje (stacking), ki je bil pri meni samo v obliki povprečne napovedi obeh modelov.

Potrebne datoteke in za ponovitev oddaje:

- tekmovanje.py
- lpputils.py
- train.csv
- test.csv
- vreme.tsv