Miha Benčina (63160061) **Naloga 3** 19. november 2018

Uvod. Cilj domače naloge je bil iz učnih podatkov o prihodih avtobusov sestaviti čim boljši model za napovedovanje prihodov avtobusov na testnih podatkih.

Ocenjevanje točnosti. Točnost sem na učnih podatkih ocenjeval z metodo k-fold cross-validation in sicer 10-fold. V moji kodi se to izvaja v funkciji crossvalidate, kjer v for zanki vsakič "izločimo" $\frac{1}{10}$ učnih podatkov (torej najprej prvo desetino, nato drugo, ...), ki jih bomo kasneje uporabili za testne. Iz ostalih $\frac{9}{10}$ podatkov pa zgradimo napovedni model s funkcijo zgradiModel. Nato model testiramo na testih, ki smo jih prej izločili. Napoved je povprečje razdalj vseh testiranj.

Napovedni modeli.

ure na učnih podatkih sem uporabil povprečni čas za odhode vsako uro, saj je ob nekaterih urah več zastojev na cesti

meseci namesto, da bi uporabil vse mesece, sem po številnih testiranjih vzorec skrčil le na 2 meseca in sicer na november in februar, saj sta najbolj reprezentativna

vreme uporabil sem podatke o vremenu decembra 2012. V primeru dežja ali snega, se je potovalni čas ustrezno podaljšal

imena linij namesto številk linij sem uporabil imena linij (tako so se upoštevali tudi odhodi avtobusov v garažo)

Rezultati. Vsakič, ko sem na novo testiral sem dodal kakšno novo funkcionalnost, tako se funkcionalnosti v tabeli 3 spreminjajo oziroma dodajajo.

Tabela 1: Prikazuje rezultate naloge

Ime metode	Oddaja	Lokalna ocena	Tekmovalna ocena
povprečje vseh	2018-11-15 14:18:19	489.31	531.38
meseci	2018-11-15 13:50:34	431.09	517.51
povprečje linij	2018-11-15 19:18:11	255.45	298.19
imena	2018-11-15 20:20:02	177.82	186.20
vreme*	2018-11-19 16:24:43	177.47	186.11

Komentarji na rezultate:
povprečje vseh: V tej metodi sem le izračunal povprečje vseh linij in ga prištel času odhodov.
meseci: To je bil moj prvi poizkus, ki je bi boljši od drugega, ki sem ga opisal zgoraj. Je skoraj enak kot prvi (povprečje vseh), s to izjemo, da sem uporabil le november in februar, tako kot v predtekmovanju.
povprečje linij: Tu namesto povprečja vseh ločil linije po številkah, prav tako pa sem uporabil povprečje po urah, ki deluje dosti bolje, kot če odhodov po urah ne ločimo, zato je bil preskok

kar velik.
imena: Ta metoda je precej podobna zgornji, le da sem namesto številk linij uporabil imena linij (npr. STANEŽIČE - MEDVODE namesto 15). To je naredilo veliko razliko, saj na ta način upoštevamo tudi, ali gre avtobus v garažo kot tudi iz katere končne postaje je štartal.
vreme*: Moja metoda, ki predstavlja končno oddajo, ki prejšnjo nadgradi z uporabo vremena, ki v primeru slabega vremena napove daljši čas potovanja avtobusa.

Dodatno (+10%). V svojo nalogo sem vključil tudi zunanji vir iz spletne strani Arso in sicer podatke o vremenu, ki so priloženi v datoteki (ljvreme_2012.csv). Tako sem izboljšal natančnost mojega modela, kot je to razvidno že zgoraj.

Izjava o izdelavi domače naloge. Domačo nalogo in pripadajoče programe sem izdelal sam.