



GDi STEM nagrada za
izvrsnost:

Primjena modela strojnog
učenja za predikciju trajanja
istovara vozila

ZADATAK

Primjena modela strojnog učenja za predikciju trajanja istovara vozila

GDi Ensemble Fleet je dio GDi portfelja softverskih rješenja koji služi za nadgledanje flote mobilne imovine poput automobila, kamiona, strojeva, plovila, ljudi ili dobara. Na terenu, resursi iz flote obilaze određene lokacije na kojima obavljaju zadatke (npr. dostava, rad po zadanom radnom nalogu i slično). Prikupljanjem telemetrijskih podataka s vozila u realnom vremenu dobiva se uvid o poziciji i kretanju vozila kroz vrijeme. Uz dostupnost tih podataka, moguće je povećati efikasnost uporabe flote smanjenjem operativnih troškova i povećanjem dostupnosti vozila kroz informiranije logističko planiranje.

Problemsko područje

Jedno od mogućih područja primjene spomenutih rješenja jest logistika opskrbe poslovnica u prehrambenoj industriji. U ovome zadatku, problemsko područje je opskrba prodajnih mjesta. Svako prodajno mjesto (poslovnica) nalazi se na određenoj lokaciji i ima definirane vremenske prozore unutar kojih mora biti opslužena u zadanome danu. Vremenski prozori, odnosno njihovo trajanje mogu biti različiti za različite poslovnice, ali i različiti za istu poslovnicu na različite dane. Prodajna mjesta opslužuju dostavni kamioni koji ih obilaze kroz dan; kamion u neko vrijeme pristupa poslovnici te vozač istovaruje kašete i kutije s proizvodima prema narudžbama za tu poslovnicu. Pritom vrijedi da se u određenu poslovnicu može dostavljati jedna ili više narudžbi, gdje svaka narudžba sadrži određen broj proizvoda (iste ili različite vrste). Ukupni volumen i težina proizvoda za dostavu variraju od narudžbe do narudžbe, a tako i ukupni volumen i težina istovara variraju od poslovnice do poslovnice. To utječe na ukupno vrijeme koje vozilo provede u istovaru pri nekoj poslovnici uz razne druge parametre, npr. o dostavljaču (vozilu), vremenu u danu, itd. Uzimanje u obzir realnije procjene trajanja istovara kod svake pojedine poslovnice moglo bi poboljšati planiranje cjelokupne dostave i povećati efikasnost uporabe flote vozila.

Zadatak

[Na ovoj poveznici nalazi](#) se zadani podatkovni skup iz opisanog problemskog područja. Radi se podskupu stvarnih (anonimiziranih) podataka koji su preuzeti iz *GDi Ensemble* sustavâ.

Vaš je glavni zadatak uz pomoć zadanog podatkovnog skupa implementirati rješenje za predikciju vremena istovara. Detalji zadatka su opisani u daljnjem tekstu.

Opis podatkovnog skupa

Zadani podatkovni skup sadrži podatke o dostavi 400 narudžbi na određene lokacije (poslovnice) flotom od 17 vozila u rasponu od 10 dana (*date*). Jedan zapis u podatkovnom skupu odnosi se na dostavu jedne narudžbe (*orderID*), pri čemu je ta narudžba dostavljena jednim vozilom (*vehicleID*) na neku lokaciju (*locationID*). (*Napomena*: u slučaju istovara više narudžbi ($n > 1$) u sklopu jedne te iste dostave nekim

vozilom, u podatkovnom skupu će se nalaziti n zapisa). Također, moguće je da se u jednom danu u neku poslovnicu dostave obavljaju u više navrata (više puta), pri čemu dostavno vozilo može svaki puta biti različito.

U svakom zapisu bilježi se zadani vremenski prozor (engl. *time window*) koji vrijedi za dostavu te narudžbe u toj poslovnici: dostava bi se trebala obaviti u nekom vremenskom pod-intervalu zadanog vremenskog prozora (dakle, unutar njega). Zadani vremenski prozor je definiran vremenom početka i završetka (*scheduledDeliveryTWStart*, odnosno *scheduledDeliveryTWEnd*), a redundantno je navedeno i njegovo trajanje (*scheduledDeliveryTWDuration*).

Pored zadanog vremenskog prozora, bilježi se i stvarni vremenski prozor trajanja dostave (onako kako se dostava stvarno dogodila) te, redundantno, i njegovo trajanje (*realDeliveryServiceDuration*). Slično kao i zadani vremenski prozor za dostavu, stvarni vremenski prozor dostave zabilježen je vremenom početka (*realDeliveryServiceStart*) i završetka (*realDeliveryServiceEnd*).

U svakom od zapisa o narudžbi navedene su i sljedeće pojedinosti narudžbe: ukupan volumen narudžbe (*totalOrderVolume*), ukupna težina narudžbe (*totalOrderWeight*), ukupan broj proizvoda sadržanih u narudžbi (*totalOrderQuantity*) te ukupan broj različitih proizvoda sadržanih u narudžbi (*totalOrderDistinctQuantity*).

U tablici 1. opisana su značenja pojedinih podataka u podatkovnom skupu.

Tablica 1 Opis podatkovnog skupa

naziv	značenje
date	datum (i narudžbe i dostave)
vehicleID	identifikator vozila koje je obavilo dostavu
locationID	identifikator lokacije (poslovnice) na kojoj se obavila dostava
orderID	identifikator dostavljene narudžbe
scheduledDeliveryTWStart	početak zadanog vremenskog prozora za dostavu
scheduledDeliveryTWEnd	završetak zadanog vremenskog prozora za dostavu
scheduledDeliveryTWDuration	trajanje zadanog vremenskog prozora za dostavu
realDeliveryServiceStart	početak stvarnog vremenskog prozora dostave
realDeliveryServiceEnd	završetak stvarnog vremenskog prozora dostave
realDeliveryServiceDuration	trajanje stvarnog vremenskog prozora dostave
totalOrderVolume	ukupan volumen narudžbe
totalOrderWeight	ukupna težina narudžbe
totalOrderQuantity	ukupan broj proizvoda sadržanih u narudžbi
totalOrderDistinctQuantity	ukupan broj različitih proizvoda sadržanih u narudžbi

Opis zadatka

Potrebno je implementirati programsko rješenje za predikciju vremena istovara i za evaluaciju implementiranog modela, te je potrebno izraditi dokumentaciju Vašeg rješenja.

Vaš je zadatak:

- Implementirati programsko rješenje za predikciju vremena istovara (regresija).
- Implementirati programsko rješenje za evaluaciju modela pomoću peterostruke [cross-validacije](#). Kao metriku upotrijebiti RMSE (Root Mean Squared Error).
- Dokumentirati rješenje. Dokumentacija treba sadržavati:
 - opis rješenja
 - jasno istaknute rezultate peterostruke *cross*-validacije u metrici RMSE
 - upute za pokretanje programskog koda.

Dodatne napomene:

- U zadanom podatkovnom skupu dio varijabli čine kategoričke varijable, a dio čine kontinuirane. Pokušajte u modelu iskoristiti obje vrste varijabli.
- Obratite pažnju na to koje su varijable važne. Primjerice, ne vrijedi nužno da ukupna težina i volumen određuju trajanje istovara; u slučaju kada se dostavlja mali broj kašeta, dostavljač ih može naslagati jedne na drugu i sve ih odjednom prenijeti iz vozila u poslovnici.
- Dozvoljeno je korištenje tehnologija po vlastitom izboru. Naša je preporuka korištenje *data science* alata za programski jezik Python (pandas, scikit-learn, itd.).
- Svoja programska rješenja postavite na *git*, te poveznicu na *git* repozitorij uključite u dokumentaciju. Moguće je predati i *Jupyter notebook*.

Pravila

Prijava, sudjelovanje i predaja rješenja

Pravilnik o kriterijima i postupku dodjele priložen je raspisu natječaja. Postupak prijave i rok za predaju rješenja navedeni su u pravilniku.

Sve prijavljene natjecateljice mogu do zadanog roka dostaviti svoje rješenje zadatka putem e-maila na gdi.hr@gdi.net, s predmetom poruke „GDi STEM nagrada 2021 – konačno rješenje“. Sve dodatne e-mailove slati na iste adrese, s predmetom poruke „GDi STEM nagrada 2021 – vaše osobno ime“.



Kriteriji bodovanja

Odabir najboljeg rješenja odvija se u dva kruga. Dio natjecateljica koje zadovolje kriterije prvog kruga ući će u uži izbor u drugom drugu.

- U prvom krugu, predana rješenja se boduju na sljedeći način:
 - 40% za rezultat evaluacije modela pomoću peterostruke *cross*-validacije u metrici RMSE. Provjeravat će ispravnost testiranja modela (nikako ne testirati rješenje na skupu za treniranje 😊).
 - 30% za kvalitetu predanog programskog koda
 - 30% za kvalitetu i razumljivost dokumentacije.
- U drugom krugu natjecateljice će prezentirati svoje rješenje te će prezentacija činiti finalnu komponentu bodovanja.

Rezultati natječaja će biti objavljeni početkom studenog 2021., a nagrade za najbolja rješenja će biti dodijeljene na Danu Fakulteta FER-a.