



# GDi for STEM

Nagrada za izvrsnost

Primjena modela strojnog učenja za  
predviđanje potražnje (Demand  
Forecasting)

ZADATAK



## Uvod

U komercijalnim industrijama koje se bave proizvodnjom i/ili isporukom različitih tipova proizvoda potrebno je provoditi kratkoročno i dugoročno planiranje potrebnih resursa kako bi se isporuka mogla neometano odvijati. Za planiranje potrebnih resursa često se provodi predviđanje potražnje iz povijesnih podataka (dnevne količine isporučenih proizvoda). U ovome zadatku fokusirat ćemo se na statističku analizu povijesnih podataka i testiranje modela za predikciju baziranih na strojnom učenju.

## Zadatak

Na ovoj poveznici: <https://github.com/intgdi2021/gdi-for-stem-2022> nalazi se zadani podatkovni skup kojega je potrebno analizirati. Radi se o podskupu stvarnih (anonimiziranih) podataka koji su preuzeti iz **GDi Ensemble** sustava.

Vaš je glavni zadatak napraviti analizu podatkovnog skupa te odabrati i testirati model za predviđanje potražnje. Detalji zadatka su opisani u daljnjem tekstu.

## Opis podatkovnog skupa

Skup podataka sastoji se od tablice proizvoda s naručenom količinom za svaki dan kada je taj proizvod bio naručivan. *Za dane kada je naručena količina 0, redak u tablici ne postoji.*

Podaci koji se nalaze u skupu su iz perioda od 26.6.2021. do 19.7.2022.

U skupu se nalaze podaci za 45 različitih vrsta proizvoda, a svaki je označen jedinstvenom ID oznakom.

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

*Tablica 1 Opis podatkovnog skupa*

Naziv	Značenje
Date	datum narudžbe
ItemId	identifikator vrste naručenog proizvoda
Quantity	količina naručenog proizvoda za taj dan

## Opis zadatka

Potrebno je napraviti statističku analizu podatkovnog skupa, implementirati programsko rješenje za predikciju potražnje proizvoda iz danog skupa, evaluirati implementirani model te izraditi dokumentaciju Vašeg rješenja.

Vaš je zadatak:

- Napraviti statističku analizu podatkovnog skupa
  - Izračunati standardne statističke mjere, kreativne analize i vizualizacije će biti dodatno nagrađene
- Implementirati programsko rješenje za predikciju buduće potražnje proizvoda tako da se predviđa količina naručenih proizvoda po vrsti na razini dana.
- Implementirati programsko rješenje za evaluaciju modela. Kao metriku upotrijebiti RMSE (Root Mean Squared Error). Validacija se pokreće nad zadnjih 30 dana, tj. U periodu od 20.6.2022. do 19.7.2022. (uključivo). Podaci za validaciju moraju biti izdvojeni iz skupa za treniranje.
- Dokumentirati rješenje. Dokumentacija treba sadržavati:
  - opis rješenja
  - jasno istaknute rezultate validacije u metrici RMSE
  - upute za pokretanje programskog koda.

*Dodatne napomene:*

- 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 adresu [gdi.hr@gdi.net](mailto:gdi.hr@gdi.net), s predmetom poruke „GDİ STEM nagrada 2022 – konačno rješenje”. Sve dodatne e-maile možete slati na istu adresu, s predmetom poruke „GDİ STEM nagrada 2022 – vaše osobno ime”.

### Kriteriji bodovanja

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

- U prvom krugu, predana rješenja se boduju na sljedeći način:
  - 30% za statističku analizu podatkovnog skupa
  - 30% za rezultat evaluacije modela s RMSE metrikom. Provjeravat će ispravnost testiranja modela (nikako ne testirati rješenje na skupu za treniranje 😊).
  - 20% za kvalitetu predanog programskog koda
  - 20% 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 2022., a nagrade za najbolja rješenja će biti dodijeljene na Danu Fakulteta FER-a.