

Napomena: Na Desktop-u preimenovati direktorijum `mu.postavke` u oblik `mu_Prezime_Ime_Indeks` (npr. `mu_Peric_Pera_mi12082`). Unutar tog direktorijuma pokrenuti jupyter server komandom `jupyter lab` ili `jupyter notebook`.

1.
 - (a) Učitati podatke iz datoteke `online_shoppers_intention.csv`. Prikazati referentne statistike za numeričke attribute koristeći `pandas` metod `describe`.
 - (b) Vrednosti za kolone `VisitorType` i `Month` zameniti rednim brojevima počevši od 0. Dozvoljeno je koristiti `pandas` metod `replace`.
 - (c) Učitane podatke razdvojiti u dva podskupa, X gde se nalaze svi atributi bez ciljane promenljive `Revenue`, i y gde se nalazi ciljna promenljiva `Revenue`. Izvršiti `one-hot` enkodiranje kategoričkih promenljivih (koristiti odgovarajuću funkciju biblioteke `pandas`).
 - (d) Podeliti podatke na skup za obučavanje i testiranje u razmeri 3:1. Za `random_state` parametar koristiti 7 i primeniti stratifikaciju po ciljnoj promenljivoj.
 - (e) Napraviti model logističke regresije sa regularizacionim parametrom C koji ima vrednost $1e-30$ i obučiti ga na podacima za obučavanje.
 - (f) Izračunati meru tačnosti (eng. `accuracy`), prikazati klasifikacioni izveštaj nad trening i test podacima, i prikazati koeficijente modela. Da li je ovo dobar klasifikator? Obrazložiti odgovor.
 - (g) Koristeći `GridSearchCV` isprobati parametre:
 - `C`: [0.01, 0.1, 0.5, 1, 10, 200, 500, 1000],
 - `class_weight`: `balanced`,
 - `penalty`: [ℓ_1 , ℓ_2].Potom odabrati najbolji klasifikator i prikazati njegove parametre.
Za najbolji klasifikator prikazati izveštaj klasifikacije nad skupovima za obučavanje i testiranje kao i tačnost (eng. `accuracy`) nad istim skupovima.
2.
 - (a) Učitati podatke za dijabetes koristeći `scikit-learn` modul `datasets` i funkciju `load_diabetes`. Na standardni izlaz ispisati informacije o tome koliko atributa i instanci skup podataka poseduje.
 - (b) Primeniti algoritam analize glavnih komponenti (eng. `PCA`) nad podacima i izvršiti smanjenje dimenzionalnosti na 6 komponenti (odnosno dimenzija). Na standardni izlaz ispisati koliko varijanse je objašnjeno nakon primene algoritma.
 - (c) Podeliti skup podataka na podatke za obučavanje i testiranje u razmeri 3:1. Za `random_state` postaviti 7. Izvršiti standardizaciju podataka.
 - (d) Napraviti neuronsku mrežu sa propagacijom unapred koja se sastoji iz 100 neurona u skrivenom sloju sa `ReLU` aktivacijom. Kao optimizator postaviti `Adam` a kao funkciju greške koristiti srednje-kvadratnu grešku.
 - (e) Izvršiti obučavanje mreže nad podacima za obučavanje. Kao podatke za validaciju postaviti podatke za testiranje. Za broj epoha postaviti 400, `batch_size` postaviti na 64. Kako Vam se ne bi prepunila konzola, preporučuje se da se postavi `verbose=0`.
 - (f) Nacrtati grafikon za funkcije greške tokom obučavanja za skup podataka za obučavanje i testiranje. Dodati legendu u kojoj je označeno koja funkcija pripada kojim podacima.