### Ime, priimek:

LV02: Strojno učenje s scikit-learn

* Knjižnjica scikit-learn

# Scikit-learn knjižnjica

Namenjena je strojnemu učenju, torej učenju modelov, ki na podlagi podatkov izvajajo npr. klasifikacijo v razrede, regresijo (napoved vrednosti parametra), ali nenadzorovano učenje (rojenje - clustering).

<https://scikit-learn.org/stable/index.html>

Slika, ki vsebuje besede besedilo, posnetek zaslona, oblikovanje

Opis je samodejno ustvarjen Slika, ki vsebuje besede besedilo, posnetek zaslona, pisava, tiskanje

Opis je samodejno ustvarjen

## Podatki / Datasets

#### Preglej tri generatorje umetnih podatkov in te podatke izriši.

#### Opiši posamezni podatkovni set (razporeditev vzorcev, število razredov, število značilk – features, najpomembnejši parametri pri generiranju podatkov).

#### Kakšni so vzorci istega seta pri večkratnem generiranju ? Kako (s katerim parametrom) bi dosegel, da vedno dobiš enake podatke (vzorce)?

## Učenje in evalvacija modela

Metrike za oceno kvalitete modela : <https://scikit-learn.org/stable/modules/model_evaluation.html#classification-metrics>

#### Kaj pomenijo spremenljivke, in kako jih generiramo : X\_train, X\_test, y\_train, y\_test

#### Pomen in vpliv parametrov: test\_size, random\_state ?

#### Izvedi učenje izbranega modela in vstavi rezultate:

#### Izpiši y\_pred. Kaj predstavlja?

#### Dodatno: izračunaj delež, koliko napovedi je pravilnih (napiši metodo)

#### Vstavi in interpretiraj pomen matrike Confusion matrix.

#### Razloži pomen metrik: precision, recall, f1-score, accuracy ( support )

#### Preskusi print\_stats metodo. Komentiraj Precision-Recall in ROC krivulje.

#### Primerjaj vsaj 3 ML modele, prikaži njihove rezultate in jih komentiraj (glede na zahtevnost izbranega testnega seta podatkov)

#### Izbrani model poženi večkrat, pri čemer mu nastavljaj parametre na različne vrednosti. Vstavi in komentiraj rezultate.

## Evalvacija s križno validacijo (Cross-validation)

#### Kako poteka K-Fold križna validacija, kaj pomeni n\_splits ?

#### Vstavi in komentiraj ROC krivuljo.

## Dodatne naloge

#### Vizualizacija napovedanih točk :

Napiši funkcijo plot\_test\_points(), ki izriše testne podatke (točke) na DecisionBoundary, vendar naj pravilno klasificirane točke prikaže kot kroge, nepravilno pa kot križce.

#### Primerjava modelov z vizualizacijo:

Generiraj sliko (matplotlib, z subplot generiraj posamezne grafe), ki za vse testne sete izriše Decision boundary in testne točke za vsak posamezni model (ena vrstica grafov za en testni set).

#### Izvedi optimizacijo parametrov izbranega modela z GridSearchCV:

Napiši kodo, ki optimizira parametre modela, za najboljši rezultat izbrane evalvacije.

#### Uporabi drug (realni) podatkovni set in primerjaj modele:

Uporabi realne podatke izbranega dataseta, za klasifikacijo ciljne spremenljivke. Primerjaj uspešnost izbranih modelov.