

Componenta proiectului în format electronic:

1. Fișier word/pdf cu prezentarea proiectului
2. Fișiere csv cu tabelele de observații input
3. Proiectul Python cu implementarea

Fișierul word/pdf cu prezentarea proiectului va fi structurat astfel:

1. Titlul proiectului. Exemplu: *Clasificarea țărilor din Europa după activitatea agricolă* (Tema 3 - Analiză de clusteri);
2. Sursa datelor (link sau specificarea explicită a sursei);
3. Specificarea numelor fișierelor de intrare csv;
4. Tabele, grafice și comentarii în conformitate cu tema aleasă.

Tematica pentru proiecte este următoarea:

Tema 1. Reducerea dimensionalității

Metode: Analiza în componente principale și Analiză factorială

Tema 2. Analiză discriminantă - Clasificare supervizată

Metode: Analiza liniară discriminantă și Discriminarea Bayesiană

Tema 3. Analiză de cluster - Clasificare nesupervizată

Metode: Algoritmi ierarhici

Pentru proiect se va alege **una** dintre aceste teme. Proiectul poate fi realizat în echipe de maxim 3 studenți.

Cerințe pentru fiecare temă:**Tema 1.***Analiza în componente principale*

A. *Tabele*: distribuția varianței, scorurile, corelațiile dintre variabilele observate și componentele principale, contribuțiile, cosinusurile, comunalitățile

B. *Grafice*: plot varianță componente cu evidențierea criteriilor de selecție a componentelor semnificative, plot scoruri, plot corelații dintre variabilele observate și componente (cercul corelațiilor), corelogramă corelații dintre variabilele observate și componente, corelogramă comunalități

Opțional: hartă scoruri

Analiza factorială

A. Aplicare test Bartlett de sfericitate și validarea modelului (modelul este respins dacă nu trece de testul Bartlett)

B. *Tabele*: Indecși KMO, varianță factori, corelații între variabilele observate și factori (corelații factoriale) cu și fără rotație, scoruri factoriale cu și fără rotație, comunalități

C. *Grafice*: Corelogramă indecși KMO, plot și corelogramă corelații factoriale cu și fără rotație, plot scoruri cu și fără rotație, corelogramă comunalități

Opțional: hartă scoruri factoriale

Tema 2

Analiză liniară discriminantă

Tabele: Puterea de discriminare a variabilelor predictor și testele Fisher pentru predictor, scoruri discriminante (variabile discriminante), puterea de discriminare a variabilelor discriminante și testele Fisher aplicate variabilelor discriminante, clasificarea în setul de antrenament, acuratețea clasificării, matricea de confuzie

Grafice: plot instanțe pe grupe și centrii în axele discriminante, distribuții axe discriminante

Discriminarea Bayesiană

Tabele: Clasificarea în setul de antrenament, acuratețea clasificării, matricea de confuzie
Aplicarea celui mai bun model pe setul de aplicare și salvarea predicțiilor într-un tabel. Alegerea celui mai bun model se face după acuratețea globală sau indexul Cohen-Kappa.

Tema 3. Analiză de clusteri

Vor fi calculate cel puțin două partiții: partiția optimală și cel puțin încă o partiție. Partiția optimală se va determina prin algoritmul Elbow.

La fiecare partiție vor fi evidențiate:

- componența partiției (clusterii cu instanțele aparținătoare);
- scorul Silhouette și graficul Silhouette;
- graficul dendrogramă;
- plot instanțe pe clusteri în primele două axe principale (componente);
- histograme pe clusteri.

Opțional: hartă clusteri