

Tema proiectului: Detectarea anomalieiilor

Pentru proiectul de detectare a anomalieiilor, vom folosi framework-ul *PyOD*, un pachet Python destinat exclusiv pentru detectarea anomalieiilor. Scopul acestui proiect este de a construi și evalua un sistem de detectare a anomalieiilor pentru un set de date real, utilizând mai multe metode disponibile în PyOD. Acest proiect va dezvolta abilități esențiale în preprocesarea datelor, aplicarea unor modele diverse de detectie a anomalieiilor și evaluarea performanței acestora.

Cerințele proiectului

1. **Alegeti un set de date ce se preteaza pe problema de detectie a anomalieiilor.**
(Vezi referinta 2 din link-uri).
2. **Preprocesarea și explorarea datelor**
 - Curățați datele de zgomot sau datele lipsă, după cum este necesar.
 - Explorați setul de date utilizând metode de analiză statistică și vizualizări. Analizați distribuția variabilelor pentru a înțelege tiparele generale și posibilele anomalii.
 - Normalizați datele dacă este cazul, mai ales dacă folosiți metode care necesită date standardizate.
3. **Alegerea și aplicarea metodelor de detectare a anomalieiilor**
 - Utilizați framework-ul *PyOD* pentru a aplica cel puțin **2 metode** diferite de detectie a anomalieiilor. De exemplu:
 - Metode bazate pe distanță, cum ar fi KNN (K-Nearest Neighbors) sau One class SVM.
 - Metode probabilistice/statistice, cum ar fi ABOD (Angle-Based Outlier Detection) sau LODA (Lightweight Online Detector of Anomalies).
 - Descrieți modul de funcționare al fiecărei metode.
4. **Implementare și ajustare**
 - Implementați fiecare metodă de detectie și ajustați parametrii pentru a obține performanțe optime.
 - Realizați un cod modular și bine structurat, utilizând framework-ul *PyOD* pentru toate modelele.
5. **Evaluarea performanței**
 - Măsurăți performanța fiecărei metode utilizând metrice de evaluare cum ar fi *Indicele Dunn, Silueta, scoruri de coerență*.
 - Compară și interpretează rezultatele pentru fiecare metodă. Explicați de ce anumite metode au performanțe mai bune decât altele pentru setul de date ales.

- Dacă este posibil, vizualizați anomaliiile detectate și comparați rezultatele fiecărei metode.

6. Analiză și concluzii

- Analizați rezultatele generale și oferiți o concluzie finală cu privire la cele mai eficiente metode pentru detecția anomalilor în setul de date ales.

7. Documentarea proiectului trebuie să conțină următoarele secțiuni:

- (0.5p) Descrierea problemei: ce fel de anomalii sunt prezente în setul de date (locale sau cluster)?
- (1.5p) Descrierea setului de date și explorarea lui (punctul 2)
- (1.5p) Metodele de detectare a anomalilor utilizate (descrierea lor)
- (1.5p) Implementarea fiecărei metode și ajustările făcute
- (2p) Evaluarea performanței și interpretarea rezultatelor.
Concluzii. Bibliografie.

Cerințe adiționale

- Utilizați Jupyter Notebook sau Google Colab pentru implementare și documentare interactivă.

Deadline și livrabile

- Data limită: 12.01.2026, ora 14:00
- Livrabile: Notebookul cu codul și explicațiile (în format .ipynb)

Link-uri utile:

1. PyOD Documentation: [PyOD Official Documentation](#)
2. Referințe pentru seturi de date: [Kaggle](#), [UCI Machine Learning Repository](#)