

Adatelemzés és adatbányászati módszerek

2. féléves feladat

Válasszon ki egy adathalmazt (kaggle.com, seaborn adatcsomag, saját adatok) (kivéve iris.csv) és végezze el az adatok elemzését Python-ban az alábbi módszertant alkalmazva.

1. Töltse be az adatokat és végezze el a szükséges adattisztítási lépéseket.
2. Elemezze az adathalmazt leíró statisztikai mutatókkal, grafikonokkal. Szűrje ki a kiugró értékeket.
3. Keresse meg a változók között fennálló korrelációs viszonyokat.
4. Klaszterezze az adatokat egy kiválasztott algoritmussal, vagy készítsen osztályozó modellt az adathalmazhoz egy kiválasztott módszerrel és mutassa meg a modell pontosságát.

Az elemzésről készítsen pdf dokumentumot, amely tartalmazza: az adathalmaz forrását és leírását, az alkalmazott eljárások rövid ismertetését, a futási eredményeket és az eredmények kiértékelését. A feladathoz csatolni kell a Python kódot is magyarázattal ellátva (kommentezve).

Leadási határidő: 2023. december 8.

A feladat leadása az aláírás megszerzésének a feltétele.

Feladat leadás módja: MS Teams feladatkiíráshoz feltölteni