

Tárgytematika / Course Description

Adatelemzés

GKLM_MSTM025**Tárgyfelelős neve /****Teacher's name:** dr. Harmati István Árpád**Félév / Semester:** 2024/25/2**Beszámolási forma /****Assesment:** Vizsga**Tárgy heti óraszám /****Teaching hours(week):** 0/0/0**Tárgy féléves óraszám /****Teaching hours(sem.):** 12/0/0

OKTATÁS CÉLJA / AIM OF THE COURSE

A tárgy célja olyan módszerek, megoldások, szoftver rendszerek ismertetése, amelyek számítógépen megvalósítható adatelemzést tesznek lehetővé. Az erre vonatkozó gépi megoldások kiterjednek az üzleti döntési folyamatok támogatására, adatfeltárás, adatkutatás automatizálására.

TANTÁRGY TARTALMA / DESCRIPTION

- Az adatelemzés ill. gépi tanulás fogalma, célja, folyamata. Néhány látványosabb alkalmazás. A gépi tanulás alapvető feladatai: Osztályozás, regresszió, klaszterezés. Matematikai alapok átlátható ismertetése.
- Python programozási alapok: A nyelv jellemzői, egyszerű adattípusok, kollekciók, vezérlési szerkezetek.
- Python programozási alapok: Comprehension-ök, kicsomagolás, haladó indexelés és iterálás, függvények, fájlkezelés.
- A NumPy numerikus számítási csomag. Tömbök létrehozása, résztömbök, műveletek, broadcasting. Egyváltozós lineáris regresszió.
- A K legközelebbi szomszéd algoritmus. Tesztelés a Római Helyszínelők feladaton.
- A pandas adatelemző csomag. Series és DataFrame adatszerkezet. CSV fájlok betöltése. Légszennyezettségi adatok elemzése.
- Többváltozós lineáris regresszió. Tesztelés a Boston Housing adathalmazon. A scikit-learn alapjai. Keresztkiértékelés.
- Logisztikus regresszió. Egy- és többváltozós logisztikus eset. Tanítás Newton-módszerrel. Tesztelés a Wisconsin Breast Cancer adathalmazon.
- A túltanulás jelensége. L1 és L2 regularizáció. Ritka mátrixok. Regularizált ritka logisztikus regresszió tesztelése az SMS Spam adathalmazon.
- Neurális hálózatok. A többrétegű perceptron modell. Tanítás sztochasztikus gradiens módszerrel, tesztelés a Phishing Websites adathalmazon.
- Döntési fák. A döntési tönk és a döntési fa modell, tanítás "brute force" módszerrel, tesztelés a Boston Housing adathalmazon.
- Ensemble módszerek. Véletlen erdő, gradient boosting. Tesztelés a Boston Housing adathalmazon.
- Klaszterezés. A K-means algoritmus. Adatvizualizáció. A t-SNE algoritmus.

SZÁMONKÉRÉSI ÉS ÉRTÉKELÉSI RENDSZERE / ASSESSMENT'S METHOD

Projektfeladat vizsgán bemutatva vagy feladatmegoldás Pythonban.

0-49%: elégtelen

50-59%: elégséges

60-74%: közepes

75-84%: jó

85-100%: jeles

KÖTELEZŐ IRODALOM / OBLIGATORY MATERIAL

<https://matplotlib.org/>

<https://seaborn.pydata.org/>

https://pandas.pydata.org/docs/getting_started/intro_tutorials/index.html

<https://plotly.com/python/>

AJÁNLOTT IRODALOM / RECOMMENDED MATERIAL

<https://matplotlib.org/>

<https://seaborn.pydata.org/>

https://pandas.pydata.org/docs/getting_started/intro_tutorials/index.html

<https://plotly.com/python/>