

Üzleti Elemzések Módszertana

Tantárgyi útmutató

Kuknyó Dániel
Budapesti Gazdasági Egyetem

2023/24
2.félév

Elérhetőségek

- E-mail: daniel.kuknyo@mailbox.org
- Messenger: Dani Kuknyo
- A tantárgy Git tárhelye: [basictask/elemzesmodszertan](https://github.com/basictask/elemzesmodszertan)
- Kérek mindenkit, Teams és Coospace felületen ne írjon, mert nem olvasom rendszeresen.
- Bármilyen kérdéssel és problémával keressetek nyugodtan.

A tantárgy tematikája

- 1 Regressziós eljárások
- 2 Osztályozási eljárások
- 3 Regularizált lineáris modellek
- 4 Döntési fák
- 5 Együttes tanuló algoritmusok
- 6 Tartó vektor gépek
- 7 Felügyelet nélküli tanulás
- 8 Generatív modellezés
- 9 Ajánló rendszerek
- 10 Neurális hálózatok
- 11 Megerősítéses tanulás

A jegy összetétele

- Összesen 100p gyűjthető
- Két elméleti teszt, egyenként 15-15p: az előadáson elhangzottakból
- 70p csoportos projekt feladat:
 - Max. 3 fős csapatok
 - Téma kiválasztása és kidolgozása
 - A projekt összetétele:
 - Jelenléti előadás
 - Elemzési csővezeték (program)
 - Dokumentáció
- Az órákon a jelenlét kötelező.

Ponthatárok

- A teljes pontszám az elméleti tesztek és a beadandók összege
- Ponthatárok:
 - $90 \leq x \leq 100 \rightarrow 5$
 - $80 \leq x < 90 \rightarrow 4$
 - $70 \leq x < 80 \rightarrow 3$
 - $60 \leq x < 70 \rightarrow 2$
 - $x < 60 \rightarrow 1$
- Ha a Hallgatónak nem sikerül elérnie a minimum követelményt, tehet vizsgát.

A téma kidolgozása

- A témáktól el lehet térni olyan algoritmusokra, modellfajtákra, amelyek az óra tárgyát nem képezték. Ha pl. valaki be szeretne mutatni egy transfertanulási esettanulmányt, teljes értékű munkának számít! (Természetesen a módszertani témán belül maradva)
- **A munka egyedisége és eredetisége elvárt és ellenőrzött.**
- Konzultációs időre lehetősége van minden csapatnak, ahol feltehetik a kérdéseiket, tanácsot kérhetnek. Ez kb. 20-30 perc, Teams felületen.
- Ha nem teljesen világos a projekt terjedelme (mit, mennyit, milyen részletesen...) konzultáció keretein belül közösen segítünk meghatározni.

A téma kidolgozása

- Terjedelem: min. 4-5 oldal képek nélkül.
- Tartalma:
 - Az elemzés célja, megválaszolandó kérdések
 - Adatforrások bemutatása, tisztítási módszerek és ezek elméleti alapjai
 - A felhasznált módszerek és ezek elméleti alapjai
 - A hiperparaméter optimalizálás módszerei és eredményei
 - A kutatási eredmények ismertetése, vizuális bemutatása
 - Javaslatok és további fejlesztés lehetőségei, más rendszerekkel való kapcsolatok
- Csatolni kell:
 - A felhasznált adatokat, és a forrást, ahonnan az adatok származnak [.csv, .xlsx]
 - A programkódot, ami az elemzést megvalósítja [.py, .ipynb stb...]
 - Az elemzéshez tartozó dokumentációt [.pdf]
 - Az elemzésről szóló bemutatót [.pptx stb...]

Választható témák

- ➊ **Regresszió** [Lineáris, Logisztikus] és Gradiens ereszkedés [Kötegelt, Sztochasztikus, Mini-kötegelt]
- ➋ **Döntési fák, Erdő modellek** [Bagging, Pasting, optimalizálás] Turbózás, együttes tanulás [Adaboost, GBM, XGBoost]
- ➌ **Regularizált modellek** [Lasso, Ridge, Elasztikus hálók]
- ➍ **Tartó vektor gépek** [Lineáris, Polinomikus, Gauss-i]
- ➎ **Generatív modellek** [Naive Bayes, Gauss-i keverékek]
- ➏ **Ajánló rendszerek** [Kollaboratív, metaadat-alapú]
- ➐ **Dimenziócsökkentés, Klaszterezés** [PCA, K-Means, Dbscan, Gauss-i]
- ➑ **Neurális hálózatok, Mélytanulás**
- ➒ **Megerősítéses tanulás**

Az órai környezet telepítése

A csomagok és függőségek a `requirements.txt` fájlban vannak eltárolva. A csomagokat telepítéséhez az Anaconda terminálban a következő parancsot kell lefuttatni:

```
pip install -r requirements.txt
```

Ezzel az összes, a kurzus alatt vizsgált munkafüzetet futtatni lehet.