

Az Adattudomány MSc képzés záróvizsga követelményrendszere

Érvényes a tanulmányaikat 2023. szeptemberben, vagy azt követően megkezdő hallgatókra

A záróvizsga két részből áll:

1. a diplomadolgozat megvédéséből, amely magában foglalja az eredmények prezentációját (15 perc); valamint
2. az arra épülő szakmai vitából és szakmai párbeszédéből (20 perc).

A záróvizsga-bizottság a szakmai vita és párbeszéd során a jelölt szakmai felkészültségének általános értékelése céljából olyan kérdéseket is feltesz, amelyek nem kötődnek szorosan a diplomadolgozat témájához és bírálatához, de kapcsolódnak a képzés során tanult szakmai ismeretekhez.

A bizottság a szakmai vita és párbeszéd keretében elsősorban az alábbi, a záróvizsgázóktól elvárt legfontosabb szakmai kompetenciák - ismeretek és készségek - meglétét térképezi fel:

Az alábbi ismeretkörök a legszűkebb törzsanyagot ölelik fel.

I. Haladó adatbázis-kezelő rendszerek:

1. A hallgató érti a nagy teljesítményű adattárolási technológiák alapelveit: oszloptár, particionálás, memória-optimalizált adattárolás
2. A hallgató tisztában van a replikáció és a szerver oldali üzleti logika megvalósításának módszereivel, alkalmazásával
3. A hallgató a relációs struktúrán kívül ismeri a dokumentum-tárak és a gráf-adatbázisok felépítését is, és a korszerű felhő adatbázis technológiák működését és alapelveit is.

Ajánlott irodalom:

- Haladó adatbázis-kezelő rendszerek jegyzet, Moodle rendszerben

II. Adatvizualizáció és üzleti intelligencia:

1. A hallgató tisztában van az üzleti intelligencia rendszerek céljával, az adatok feldolgozásának általános folyamatával az adatforrások modellezésétől kezdve az adatformázáson át a különféle vizualizációs módszerekig.
2. A hallgató ismeri az egyszerű adatvizualizációs technikák működését, a mögöttes üzleti logikát.

Ajánlott irodalom:

- Az Adatvizualizációs és üzleti intelligencia rendszerek tananyagai a Moodle rendszerben
- Üzleti intelligencia szoftverek hivatalos oldalai.
<https://www.ibm.com/products/cognos-analytics>
<https://powerbi.microsoft.com/>
- Adatvizualizációs programkönyvtárak hivatalos oldalai.
<https://matplotlib.org/>
<https://plotly.com/>
<https://seaborn.pydata.org/>

III. Adataalapú gépi tanulás:

1. A hallgató ismeri az adataalapú gépi tanulás fő feladatosztályait, s képes egy tetszőleges problémához kiválasztani a megfelelő tanulási módszertant. A hallgató ismeri a KDD folyamat lépéseit, tisztában van az adatelőkészítés fő feladataival, ismeri a gyakorlatban alkalmazott adatelőkészítési technikákat és azok jelentőségét.
2. A hallgató ismeri a feltáró adatelőkészítés (EDA) és dimenziócsökkentés eszközeit. A hallgató ismeri az összefüggésvizsgálatok különféle módszereit és alkalmazási körét.
3. A hallgató széleskörűen ismeri a különféle alapszintű és embedded osztályozó modellek felépítését, alkalmazási korlátait és a felépítésükhöz használt különféle algoritmusok működését. A hallgató ismeri az osztályozó modellek kiértékelési módszereit.
4. A hallgató ismeri a regressziós feladatok megoldásához használt különböző regressziós modellek felépítését és felépítésükhöz használt algoritmusok működését. A hallgató ismeri a regressziós modellek kiértékelési módszereit.
5. A hallgató ismeri az osztályozó és regressziós modellek felépítésével kapcsolatos hiperparaméter hangolási technikákat, ismeri a különféle jellemzőszelekciós módszereket, valamint ismeri a kiegyensúlyozatlan adathalmazokból fakadó problémákat és a problémák megoldásához használt módszereket.
6. A hallgató ismeri a csoportosító algoritmusok működését, azok alkalmazásának előnyeit és hátrányait, valamint ismeri a csoportosítások eredményének kiértékelési módszertanát. A hallgató ismeri az outlierok feltárásához alkalmazott főbb algoritmusokat és működési elvüket.
7. A hallgató ismeri az asszociációs szabályok bányászatához használt algoritmusokat és működési elvüket, valamint az asszociációk szabályok kiértékeléséhez használt metrikákat.
8. A hallgató tisztában van a folyamatbányászat alapjaival, a bemeneti eseménynaplók formai és tartalmi követelményeivel. Ismeri a fő folyamatbányászati feladatköröket és az információk kinyeréséhez használt főbb algoritmusokat.

Ajánlott irodalom:

- A Felügyelt gépi tanulás c. kurzus diasorozata a Moodle rendszerben

- A Nem felügyelt gépi tanulás c. kurzus diasorozata a Moodle rendszerben
- Jiawei Han, Micheline Kamber. Adatbányászat, Panem, 2004
- Wil M van der Aalst. Process Mining: Data science in action. Springer, 2018

IV. Neurális hálózatok:

1. A hallgató ismeri a neurális hálózatok működését és a mögöttes matematikai számítási elveket, valamint ismeri a neurális hálózatok főbb típusait. Ismeri és érti a neurális hálózatokkal kapcsolatos alapvető fogalmakat és kapcsolódó technológiákat. Az adatalapú gépi tanulási problémakörhöz kapcsolódóan képes meghatározni a probléma megoldásához illeszkedő neurális hálózat felépítését.
2. A hallgató ismeri és érti a neurális hálózatok tanításával és finomhangolásával kapcsolatos feladatokat és módszereket. Ismeri és érti a neurális hálózatok paramétereit és a tanítás során használt hiperparamétereket, tisztában van ezek jelentőségével és a tanítási folyamatra gyakorolt hatásával. Ismeri a túltanulás és alultanulás jelenségét, valamint ismeri azon technikákat, amelyekkel ezen problémák elkerülhetők. Ismeri és érti a tréning, validációs és teszt halmazok jelentőségét, és kialakításának technikáit.
3. A hallgató ismeri a mély neurális hálózatok működését, ismeri a visszacsatolt, az autoencoder, a generatív és a mély megerősítéses tanuláson alapuló neurális hálózatok működési elvét és alkalmazási területeit.

Ajánlott irodalom:

- Mesterséges intelligencia c. kurzus diasorozata a Moodle rendszerben
- A Felügyelt gépi tanulás c. tantárgy diasorozata a Moodle rendszerben
- Mélytanuló rendszerek c. tantárgy diasorozatai és notebook-jai a Moodle rendszerben
- Stuart J. Russel and Peter Norvig: Artificial intelligence

V. Információbiztonság és adatvédelem:

1. A hallgató ismeri az adatvédelem és az információbiztonság alapelveit, fogalmait, az információs alapjogok rendszerét, és ezek kapcsolódását, érintettek jogait és kötelezettségeiket. Ismeri az adatműveletekre vonatkozó szabályokat, kockázatokat és képes a kezelt vagy feldolgozott adat vonatkozásában az információbiztonság alapelveinek megfelelő technológiákat és módszereket alkalmazni.

Ajánlott irodalom:

- Információbiztonság és adatvédelem c. kurzus diasorozata a Moodle rendszerben
- Általános adatvédelmi rendelet(GDPR), Infotv.
- HVG kiadó: Adatvédelem a gyakorlatban
- Gyaraki Réka: Az információbiztonság alapjai (2023)

A záróvizsga eredménye:

A záróvizsga érdemjegyét (ZE) a záróvizsga-bizottság állapítja meg a témavezető és bíráló által a diplomadolgozatra javasolt érdemjegyek, a szóbeli védés és az azt követő szakmai vita és párbeszéd alapján.