



Adattudomány MSc (2026) – záróvizsga tételek

Az adattudomány elméleti háttéréhez kapcsolódó alapismeretek:

1. Adatmenedzsment adattudományi projekteknél: adatélelciklus; adatok gyűjtése, előkészítése, tárolása és publikálása; adatforrások integrációja és adatfeldolgozási folyamatok; adatminőség, -profilozás és -validáció; metaadatkezelés; reprodukálható adatelőkészítés adattudományi feladatokban.
2. Adatmodellezés és adattárolási megközelítések: koncepcionális, logikai és fizikai adatmodellezés; ER-modell; relációs modell, normalizálás és denormalizálás; haladó SQL-alapú adatkezelés; adattárház-modellezés és elemzési célú adattárolás; félig strukturált és nem relációs adatmodellek; adatmodellezési szempontok elemzési és gépi tanulási feladatokban.
3. Gépi tanulás: felügyelt és nem felügyelt tanulás; egy- és többváltozós lineáris regresszió; gradienstsökkentési eljárás; leírók normálása; polinomiális regresszió; normál egyenlet; logisztikus regresszió; két- és többosztályos osztályozás; regularizáció; alul- és túltanulás; regularizált lineáris és logisztikus regresszió; neurális hálók; visszaterjesztési algoritmus; numerikus gradienstenőrzés.
4. Gépi tanulás: tanító-, teszt- és validációs adatfelbontás; tanítási diagnosztika; tanulási görbék; hibamérés és kiegyenlített osztályok; támasztóvektor-gépek és magfüggvények; klaszterezés; klaszterszám meghatározása; dimenziócsökkentés; anomáliadetektálás; normális eloszlás; ajánlórendszerek; tartalomalapú ajánlás; kollaboratív szűrés; elosztott feldolgozás.
5. Kísérlettervezés: oksági és nem oksági kapcsolatok megkülönböztetése; beavatkozások, döntések és változások hatásának értelmezése; randomizálás; kontrollcsoportok és összehasonlítható csoportok kialakítása; mintavétel; torzítások kezelése; kísérleti elrendezések bemutatása szakterületi példákban.
6. Oksági következtetés és hatásbecslés: megfigyeléses adatokból történő következtetés; természetes és kvázi-kísérleti helyzetek; különbségek különbsége módszer (difference-in-differences); regressziós megszakítási elrendezés (regression discontinuity); instrumentális változók; oksági gráfok; módszerek bemutatása szakterületi példákban.
7. Az adattudomány statisztikai alapjai: többdimenziós minta és jellemzői; többdimenziós normális eloszlás; főkomponens-analízis; faktoranalízis; kanonikus korrelációanalízis; maximum likelihood és Bayes-osztályozás; lineáris és kvadratikus diszkriminálás; legközelebbi társ módszer; klaszteranalízis; többdimenziós skálázás.

Az adattudomány gyakorlati hátteréhez kapcsolódó alapismeretek:

1. Felhőalapú adatrendszerek: felhőinfrastruktúra adatintenzív rendszerekben; infrastruktúra-, platform- és szoftverszolgáltatások (IaaS, PaaS, SaaS) szerepe adattárolásban és adatfeldolgozásban; adatbázis-szolgáltatások; adattárházak; adatfolyam-kezelés; nyilvános, privát és hibrid felhők adatkezelési szempontjai.
2. Felhőalapú adatrendszerek: adattó (data lake), adattárház és lakehouse architektúrák; adatfeldolgozási folyamatok; felhőalapú adattárolás; hozzáférés-kezelés; adatbiztonság; monitorozás; skálázhatóság; költség- és teljesítményszempontok; kockázat- és mentési stratégiák.
3. Adatvizualizáció: adatvizualizáció alapfogalmai; vizuális érzékelés; adatabsztrakció; adattípusok; adatok előkészítése vizualizációhoz; feladatabsztrakció; vizualizációs tervezési folyamat.
4. Adatvizualizáció: kategorikus, numerikus, ordinális, időbeli és földrajzi adatok megjelenítése; fák, gráfok és hálózatok vizualizációja; interakció; skálázhatóság; irányítópult-készítés (dashboard); adatvezérelt történetmesélés; vizualizációs eszközök.
5. Adatetika és -irányítás: adatvédelem alapjai; személyes adatok köre; adatkezelők és adatfeldolgozók; GDPR-alapelvek; érintetti jogok; adatvédelmi incidens; információszabadság és közérdekű adatok; etikai kódexek; adatirányítási alapelvek; adatminőség, átláthatóság és elszámoltathatóság; Big Data, IoT és MI-alapú adatkezelés etikai-jogi kérdései; felelős adatfelhasználás.
6. Adatorientált alkalmazásfejlesztés: adatalapú alkalmazások fejlesztési lehetőségei; kódolás nélküli és alacsony kódolású eszközök (no-code, low-code); vizuális munkafolyamat-rendszerek; adatkapcsolatok és alkalmazásintegrációk; alkalmazásprototípusok készítése; adattudományi feladatok támogatása alkalmazásfejlesztési eszközökkel.
7. Adatorientált alkalmazásfejlesztés: adatfeldolgozási és automatizálási folyamatok tervezése; ügynök-alapú és MI-támogatott automatizálás; alkalmazásprogramozási felületek (API-k); adatok betöltése, átalakítása és továbbítása; Python-alapú és MI-asszisztált fejlesztési lehetőségek; alkalmazások bevezetése és működtetése adatintenzív környezetben.