



Adattudomány MSc (2023) – záróvizsga tételek

Az adattudomány elméleti háttéréhez kapcsolódó alapismeretek:

1. Az információ biztonság alapfogalmai (CIA hármas); A kiberbiztonság eszközei, céljai, kártékony programok, támadási technológiák; Hozzáférés-szabályozás (DAC, MAC, RBAC, ABAC, CBAC); Hozzáférés-szabályozás elosztott rendszerekben;
2. Hitelesítés, felhasználó hitelesítés; Hitelesítés elosztott rendszerekben; Nyomonkövethetőség; Biztonságos üzemeltetés és incidenskezelés; Monitor Analyze Plan Execute-Knowledge (MAPE-K); Rendelet, szabványok.
3. Felügyelt- és nem felügyelt tanulás; Egy és többváltozós lineáris regresszió; Gradiens csökkenési eljárás; Sztochasztikus és mini-batch gradiens csökkentés; Leírók normálása; Polinomiális regresszió; Normál egyenlet; Logisztikus regresszió; Két- és többosztályos osztályozás; Regularizáció (alul- és túltanulás); Regularizált lineáris és logisztikus regresszió; Neurális hálók; Backpropagation algoritmus; Numerikus gradiensellenőrzés.
4. Tanító/Teszt/Validációs adatfelbontás; Tanítási diagnosztika; Tanulási görbék (tanító adathalmaz mérete); Hibamérés és kiegyenlített osztályok; Támasztóvektor-gépek és magfüggvények alkalmazása; Klaszterezés; Klaszterek számának meghatározása; Dimenziócsökkentés; Anomáliadetektálás (normális eloszlás); Ajánlórendszerek; Tartalom alapú ajánlás; Kollaboratív szűrés; Map-Reduce és párhuzamosítás.
5. Többváltozós függvények feltétel nélküli és feltételes szélsőértéke. Gradiens módszerek, megbízhatósági tartomány, Newton-módszer, kvázi Newton-módszerek, konjugált gradiens módszer, legkisebb négyzetek módszere, sztochasztikus optimalizálás.
6. Többdimenziós minta és jellemzői, többdimenziós normális eloszlás. Főkomponens-analízis. Faktoranalízis. Kanonikus korrelációanalízis. Osztályozási módszerek (maximum likelihood és Bayes osztályozás, lineáris- és kvadratikus diszkriminálás, legközelebbi társ módszer). Többdimenziós skálázás.

Az adattudomány gyakorlati hátteréhez kapcsolódó alapismeretek:

1. A felhőinfrastruktúra leírása; Az IaaS, PaaS és SaaS megkülönböztetése; Felhőalapú számítástechnika típusai közötti különbségek bemutatása (nyilvános, privát, helyhez kötött, hibrid); A felhőkörnyezet megbízhatóságának, rendelkezésre állásának, skálázhatóságának ismertetése; A felhős alkalmazások és rendszerek költségeinek bemutatása, elemzése; A felhős környezetben a rendelkezésre állási mérőszámok ismertetése;
2. A nyilvános, a privát és a hibrid felhőalapú alkalmazások megvalósítása közötti különbségek áttekintése; Virtuális környezet létrehozása (virtuális gépek, virtuális hálózat, adattárolás); A működési költségek kezelése; A lehetséges kockázatok és katasztrófa-forgatókönyvek azonosítása; Helyszíni és külső biztonsági mentési stratégia kialakítása; A felhőalapú rendszerek felügyelete; A felhőbiztonság kezelése.
3. Az adatvizualizáció alapvető fogalmai, kialakulásának története, a vizuális érzékelés szerepe; Adatabsztrakció, adattípusok, az adatok előkészítése a vizualizációhoz; Feladatok és célok, feladatabsztrakció; A vizualizáció tervezési folyamata; Kategorikus, numerikus, ordinális, időben változó és geográfiai adatok megjelenítési lehetőségei; Fák, gráfok és hálózatok vizualizációja; Interakció, skálázhatóság, animáció, színek.
4. Nagy mennyiségű adat megjelenítési lehetőségei; Dimenziócsökkentési technikák; A vizualizáció fontossága a feltáró adatelemzésben; Dashboard készítése; Storytelling; Az adatvizualizációt segítő korszerű függvénykönyvtárak és szoftverek.
5. Az adatvédelem elmélete és fejlődése az európai jogi kultúrában; A személyes adatok köre; Adatkezelők és adatfeldolgozók; A GDPR alapelvei; Az adatalany jogai; Adatvédelmi incidens, jelentési kötelezettség, vizsgálat, a nemzeti adatvédelmi hatóságok szerepe; A személyes adatok védelméhez fűződő jog megsértésének jogkövetkezményei; ePrivacy az Európai Unióban.
6. Az etikai kódexek szerepe az adatvédelemben; Az információszabadság és a közérdekű adatok nyilvánossága; Az EUB adatvédelemhez kapcsolódó esetjoga; Az adatvédelem és az etika összefüggései; Big Data: etikai és jogi kérdések; IoT: etikai és jogi kérdések.
7. Mesterséges intelligencia (MI) programozásához kapcsolódó függvénykönyvtárak; Kernel szintű futtató és fordító környezetek; Elterjedtebb függvénykönyvtárak, eszközök (scikit-learn, NumPy, SciPy, pandas, Jupyter, Matplotlib, Dataflow, Keras, TensorFlow)..