

Záróvizsga tematika

Záróvizsga tárgy neve: Big Data és üzleti intelligencia

(Big Data algoritmusok és programozás, Adattárházak és üzleti intelligencia)

Szak: Mérnökinformatikus alapképzési szak

Specializáció: Big Data és üzleti intelligencia

1. Adatbányászat: Mi az adatbányászat és miért van szükség rá? Ismertesse a tudásfeltárás folyamatát és azon belül az adatbányászat szerepét. Mutassa be az adatbányászati feladatokat, azok jellemzőit, példákkal illusztrálva.
2. Az adatok általános jellemzése és alapvető statisztikai jellemzői: Ismertesse az adattípusokat, valamint az attribútum típusokat. Mutassa be az adatok alapvető statisztikai jellemzőit átlag, medián, módusz, kvantilisek, adatok eloszlása.
3. Adatok előfeldolgozása: Miért szükséges az adatok előfeldolgozása, illetve melyek az előfeldolgozás fő feladatai? Mit értünk adattisztításon, miért van rá szükség, milyen módszerei vannak? Mi az adatintegráció, kell figyelni a megvalósítása során? Mi az adatredukció és milyen módszereket ismer annak megvalósítására?
4. Adatok összehasonlítása, adatvizualizáció: Hogyan tudja mérni az attribútumok és az adatok hasonlóságát, illetve különbségét? Mit mér és mire használható a korreláció? Miért és mire jó az adatvizualizáció? Milyen adatvizualizációs technikákat ismer?
5. Adattárházak: Mi az adattárház? Ismertesse az adattárházak fő jellemzőit Inmon definíciója szerint. Ismertesse az ETL (Extraction, Transformation and Loading) folyamatot. Melyek az OLTP és OLAP rendszerek közötti fő különbségek? Ismertesse az adatkocka lényegét és a főbb adatkocka műveleteket. Ismertesse és hasonlítsa össze a relációs adattárolási modelleket (csillagséma, hópehelyséma, galaxis séma).
6. Osztályozás döntési fákkal: Ismertesse az osztályozás feladatát és annak folyamatát. Mutassa be részletesen, példán keresztül a döntési fa készítésének folyamatát. (döntési fa, döntési fa algoritmusok, attribútum-választási stratégiák, túlilleszkedés és alulilleszkedés, szabályok készítése a döntési fa alapján)
7. Az osztályozás pontossága: Ismertesse az osztályozó modellek pontosságának mértékeit a keveredési mátrix (confusion matrix) használatával: pontosság (accuracy), hiba arány (error rate), érzékenység (sensitivity), specifikusság (specificity), megbízhatóság (precision), F1-mérték (F1-score). Ismertesse az adathalmaz tanulóhalmazra és teszhalmazra osztásának módszereit.
8. Az osztályozás alternatív módszerei: Ismertesse röviden az osztályozás alternatív módszereit (szabályalapú osztályozók, legközelebbi szomszéd osztályozók, Bayes osztályozási módszerek, neurális hálók, tartóvektor gépek, együttes módszerek) és egy tetszőlegesen kiválasztott módszert mutasson be részletesen, példával illusztrálva. Nevezzen meg az osztályozáshoz kapcsolódó néhány gyakorlati alkalmazást.

9. Asszociációs elemzés: Ismertesse az asszociációs elemzéshez használt alapfogalmakat (elemhalmaz, k-elemhalmaz, gyakori elemhalmaz, asszociációs szabály, asszociációs szabály gyakorisága és megbízhatósága). Ismertesse a gyakori elemhalmazok bányászatának folyamatát, azon belül az apriori algoritmus lényegét. Mutassa be az asszociációs szabályok feltárásának lépéseit. Nevezzen meg az asszociációs elemzéshez kapcsolódó néhány gyakorlati alkalmazást.
10. Csoportosítás (klaszterezés): Ismertesse a csoportosítás feladatát. Mutassa be röviden a csoportosítás módszereit (felosztó módszerek, hierarchikus módszerek, sűrűség-alapú módszerek) és egy tetszőlegesen kiválasztott módszert mutasson be részletesen, példával illusztrálva. Nevezzen meg a csoportosításhoz kapcsolódó néhány gyakorlati alkalmazást.
11. Rendellenességek észlelése: Ismertesse a rendellenes/kiugró értékek jellemzőit, azok okait és észlelésének módszereit. Egy tetszőlegesen kiválasztott módszer segítségével, példán keresztül mutassa be a rendellenességek észlelését. Nevezzen meg a rendellenességek észleléséhez kapcsolódó néhány gyakorlati alkalmazást.
12. A Big Data forrásai és jellemzői: Mutasson példákat, ahol nagyon sok adat keletkezik, és ahol annak feldolgozása kihívás. Ismertesse a „Big Data 4V”-jét és a további „V”-it.
13. Hadoop: Ismertesse a Hadoop ökoszisztéma fő elemeit. Mutassa be a Hadoop elosztott fájlrendszerét (HDFS), annak architektúráját, skálázhatóságát, adattárolási megoldását. Mutassa be a Hadoop ökoszisztéma elosztott programozási modelljét (MapReduce).
14. Big Data rendszerek elemzése: Ismertesse és jellemezze a gépi tanulás folyamatának lépéseit. Ismertesse a lépéseket a legnépszerűbb machine learning package, a scikit-learn használatával is. Ismertesse a felügyelt tanulás és a nemfelügyelt tanulás lényegét és mutassa be a gépi tanulás módszereit.
15. Ajánlórendszerek: Ismertesse az ajánlórendszerek lényegét, alkalmazási területeit. Mutassa be és jellemezze az ajánlórendszerek fő típusait. Mutassa be az ajánlórendszerek készítésének lépéseit egy példán keresztül.
16. Gráf adatbázisok: Ismertesse a gráf adatok jellemzőit (csomópont-tábla, él-tábla). Jellemezze a tulajdonsággráfot (property graph) és a gráf adatokból meghatározható tulajdonságokat (utak, szomszédságok, közösségek, összefüggőség). Ismertesse a Neo4J rendszer fő funkcióit.
17. Mélytanulás (deep learning): Ismertesse a neurális hálók felépítését és tanításának módszerét. Mi jellemzi a mélyneurális hálókat? Nevezzen meg a mélyneurális hálókkal kapcsolatban néhány gyakorlati alkalmazást. Mi a TensorFlow illetve a Keras? Milyen tipikus rétegei vannak egy konvolúciós neurális hálónak?