|  |  |  |  |
| --- | --- | --- | --- |
| Ainekood IFI6222.DT | NIMETUS: Kvantitatiivne digihumanitaaria | | |
| Maht EAP 4 | Kontakttundide maht: 56 + arvestus | Õppesemester: K | Arvestus |
| Eesmärk: | Luua eeldused teadmiste ja oskuste kujunemiseks digitehnoloogia kvantitatiivsete analüüsimeetodite valdkonnas.  Luua eeldused praktiliste oskuste kujunemiseks digitaalse uurimistöö tehnikatest. | | |
| Aine lühikirjeldus:  (sh iseseisva töö sisu kirjeldus vastavuses iseseisva töö mahule) | Kirjeldava statistika meetodite omandamine praktiliste näidete abil, kasutades sisendeid erisuguses vormingus allikatest. Andmete filtreerimine ja teisendamine paremini uuritavaks muutmiseks. SQL käsklused, XML vahendid.  Statistika põhimõisted ning nende kasutamine andmete iseloomustamisel.  Erinevuse tähtsuse hindamine 2x2 andmetabeli põhjal  Andmestike erinevuste ja sarnasuste leidmine sagedustabelite hii-ruut testi abil. Selle rakendused keeleteaduses, folkloristikas, ajaloos jm.  Kirjete rühmitamine metaandmete järgi.  Tunnuste grupeerimine. Sõltuvuste leidmine tunnuste vahel. Seose modelleerimine lineaarsena. Muu kujuga seosenäiteid.  Histogrammi abil andmete jaotuse nägemine  Normaaljaotuse omadused, andmestiku jaotuse võrdlemine normaaljaotusega. Näited näitajatest, mis alluvad normaaljaotusele rohkem ja nendest, mis vähem. Keskväärtuse ja standardhälbe abil kumulatiivsete väärtuste ennustamine. Tavaliste ja erakorraliste kõikumiste avastamine tekstide sõnasageduste juures.  Keskväärtuste võrdlemine t-testi abil.  Kokkuvõtete eelnev väljaarvutamine DataCube’i abil.  Mitmesuguste analüüsimeetodite tutvustus. Masinõppe võimalused | | |
| Õpiväljundid: | Oskab kasutada kvantitatiivseid meetodeid humanitaarvaldkonna analüüsiks.  Tunneb kvantitatiivsete meetodite kasutusvaldkondi.  Oskab valida sobiva meetodi vastavalt olukorrale.  Suudab hinnata tehtavate järelduste täpsust ning seda mõjutavaid tegureid. | | |
| Hindamismeetodid: | Arvestus  Kodutööde esitlemine, seminar, kahe uurimuse koostamine  [http://minitorn.tlu.ee/~jaagup/kool/java/kursused/18/kvantdh/juht.html](http://minitorn.tlu.ee/~jaagup/kool/java/kursused/juht.html) | | |
| Õppejõud: | Jaagup Kippar | | |
| Ingliskeelne nimetus: | Quantitative Digital Humanities | | |
| Eeldusaine: |  | | |
| Kohustuslik kirjandus: | Baayen, R. H. A Practical Introduction to Statistics  http://minitorn.tlu.ee/~jaagup/kool/java/kursused/17/xmlrak/baayenCUPstats.pdf | | |
| Asenduskirjandus:  (üliõpilase poolt läbi töötatava kirjanduse loetelu, mis katab ainekursuse loengulist osa) | Ainet pole võimalik läbida ainult asenduskirjanduse alusel. Kokkulepped võimalikud põhjalikuma digihumanitaaria valdkonna uurimistöö koostajatel. | | |
| Õppetöös osalemise ja eksamile/arvestusele pääsemise nõuded | Arvestuse kirja saamiseks peab esitama ja kaitsma kõik kodutööd ja uurimistöö ning osalema seminaril. | | |
| Iseseisva töö nõuded | Enamikel tunnis läbitud teemadel tuleb kodutööna valmistada oma valitud andmete põhjal sarnaselt töötav lahendus. Täpsemad seletused ning täiendused ilmuvad semestri käigus kursuse lehele | | |
| Eksami hindamiskriteeriumid või arvestuse sooritamiseks vajalik miinimumtase | Hindamiskriteeriumid, millest hindamisel lähtutakse:  Teadmised  Arvestatud: Tunneb levinumaid kvantitatiivse analüüsi meetodeid digihumanitaaria valdkonnas ning teab nende kehtivuspiire ja kaasnevaid ohte  Oskused  Arvestatud: suudab andmestikud kohendada uurimiseks sobivaks vastavalt ettevõetud meetodile, analüüsi läbi viia ja tulemusi tõlgendada. | | |
| Informatsioon kursuse sisu kohta, kursuse jaotumine teemade kaupa sh kontakttundide ajad | Läbitavad teemad  30.01 Jooniste koostamine ggplot-paketi abil, DataCamp kursus  02.02 Joonise abil loo jutustamine, mitme kihi paigutamine joonisele  06.02 Arvulise andmekogumi kirjeldamine, normaaljaotus ja selle omadused. Standardhälve ja standardviga. Jaotuse võrdlemine normaaljaotusega. Binoomjaotus, Poissoni jaotus.  09.02 Andmekogumi keskväärtuse võrdlemine arvuga, usaldusvahemikud ja tulemuse illustreerimine. T-test  13.02 Kahe andmekogumi keskväärtuste omavaheline võrdlemine, näited mitmesuguste andmestike põhjal. 2X2 tabelist järelduste tegemine ja nende tõenäosuste arvutamine  16.02 Sagedustabelite koostamine, absoluutsed ja suhtelised sagedused, reasagedused, veerusagedused.  20.02 Andmestike võrdlemine sagedustabelite alusel, hii-ruut test  23.02 Filmiandmetest sagedusandmete leidmine ja nende võrdlemine  27.02 Regressiooniseos, väärtuste ennustamine  02.03 Regressiooniseoste näited eri valdkondade andmetes  06.03 Peakomponentide analüüs  09.03 Peakomponentide analüüsi näiteid  13.03 Faktoranalüüs, võrdlus peakomponentide analüüsiga  16.03 Filmiandmete uurimistööde vahekokkuvõte ja konsultatsioon  27.03 Analüüsi tulemuste esitamise näiteid  30.03 Filmiandmete uurimistööde esitamine  03.04 Mitmemõõtmeline skaleerimine: algoritm  06.04 Mitmemõõtmelise skaleerimise näiteid  10.04 Andmeridade/objektide klasteranalüüs  13.04 Tunnuste klasteranalüüs, näited mitme valdkonna andmetega  17.04 Hierarhiline klasteranalüüs  20.04 ANOVA  24.04 Keeletarkvara võimalused. Ülevaade vahenditest ning võimalikest uurimistööde teemadest  27.04 Keeleanalüsaatorite tutvustus, seoste leidmine andmestike sees ja vahel  04.05 Masinõppe meetodid, nende tööpõhimõtted. Lähinaabrite järgi rühmakuuluvuse ennustamine (KNN)  08.05 Keeleuurimistöö konsultatsioon  11.05 Seminar  Arvestuspäev: mõnda keerukamat meetodit kasutava keeleuurimuse tutvustamine ning täiendamine arvestuse jooksul | | |

|  |  |
| --- | --- |
| Õppeainet kureeriv üksus: | Informaatika instituut |
| Kursuseprogrammi koostaja |  |
| Allkiri: |  |
| Kuupäev: |  |

Kursuseprogramm registreeritud akadeemilises üksuses

|  |  |
| --- | --- |
| Kuupäev |  |
| Õppeassistendi nimi |  |
| Allkiri |  |