

TALLINNA TEHNIKAÜLIKOOL

Inseneriteaduskond

Virumaa kolledž

Nelli Matuzilova

**R-keele baasil masinõppe meetodite uurimine.
E-kursuse õppematerjalide koostamine**

Rakendusinfotehnoloogia õppekava lõputöö

Juhendaja: O. Dunajeva, lektor

Kohtla-Järve 2019

KOKKUVÕTE

Lõputöö eesmärgiks oli uurida masinõppe meetodeid statistilise andmetöötluse programmeerimiskeele R baasil ja saadud teadmiste põhjal õppematerjalide loomine virtuaalses õpikeskkonnas Moodle EVK0020 Andmeteadus ja masinõpe kursuse jaoks. E-kursuse iga segment peaks koosnema loengumaterjalist, mudeli näite üksikasjalikust selgitusest, teema kohta saadud teadmiste testimisest ja kodustest ülesannetest kirjeldatud teadmiste kinnistamiseks.

Lõputöö käigus autor töötas põhjalikult läbi andmete teaduse ja masinaõppe põhiprintsiipe, algoritme ja mudeleid, näiteks andmete esmane analüüs ja visualiseerimine, põhifunktsioonid ja printsiipe andmete ettevalmistamise masinõpeks, lineaarregressioon, otsustuspuud. Saadud teadmiste põhjal oli koostatud toetavad teoreetilised õppematerjalid kahes keels - eesti ja inglise keeles, järgmisele teemadele - masinõpeks andmete ettevalmistamine, lineaarne regressioon, otsustuspuud. Töö kirjutamisprotsessis autor oli tutvustatud programmeerimiskeelega R, mille baasil oli koostatud näidisülesanded masinõpe näidete ja nende üksikasjalike selgitustega, ning testid ja koduülesanded teadmiste kinnistamiseks. Virtuaalses õpikeskkonnas Moodle oli koostatud EVK0020 Andmeteadus ja masinõpe kursus valmistatud materjalidega, kus autor on saanud endas proovida õpetaja rollis.

Põhiline probleem, millega autor on kokku puutunud, on ebapiisav eesti keelne materjalide kogus, uuritud teemade valdkonnas. Suur osa materjalist, mis oli kasutatud masinõppe ja andmeanalüüsi põhimõtte uurimiseks ja õppematerjalide koostamiseks, oli inglise keeles ja see võttis palju aega, et eesti keeles tõlkida.

Tulevikus autoril on plaanis uurida ülejäänud andmete teaduse ja masinaõppe meetodeid, mis EVK0020 Andmeteadus ja masinõpe kursus hõlmab, näiteks närvivõrkude ja Bayesi meetodite põhimõtted. Saadud teadmiste põhjal koostada õppematerjalid ja näidisülesanded nimetatud e-kursuse jaoks. Lisaks ettevalmistada testid ja koduülesanded teadmiste kinnistamiseks nendel teemadel.

Lõputöö tehes sai autor palju uut teadmisi andmeteadusi ja masinõppe valdkonnas, õpis kuidas valida ja rakendada sobivaid meetodeid andmeanalüüsi ja masinõppe ülesannetelahendamiseks, ning tutvus programmeerimiskeelega R. Autor sai ka kasulik kogemus virtuaalses õpikeskkonnas Moodle töötamist õpetaja rollis.