



Kandidatutkielma

Tietojenkäsittelytieteen kandiohjelma

Oppijan kehittymisen tukeminen oppimisanalytiikalla Moodlessa

Tuomas Alanen

5.6.2025

MATEMAATTIS-LUONNONTIETEELLINEN TIEDEKUNTA
HELSINGIN YLIOPISTO

Yhteystiedot

PL 68 (Pietari Kalmin katu 5)
00014 Helsingin yliopisto

Sähköpostiosoite: info@cs.helsinki.fi
URL: <http://www.cs.helsinki.fi/>

Tiedekunta — Fakultet — Faculty		Koulutusohjelma — Utbildningsprogram — Study programme	
Matemaattis-luonnontieteellinen tiedekunta		Tietojenkäsittelytieteen kandiohjelma	
Tekijä — Författare — Author			
Tuomas Alanen			
Työn nimi — Arbetets titel — Title			
Oppijan kehittymisen tukeminen oppimisanalytiikalla Moodlessa			
Ohjaajat — Handledare — Supervisors			
Lea Kutvonen			
Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages	
Kandidutkielma	5.6.2025	20 sivua, 1 liitesivua	
Tiivistelmä — Referat — Abstract			
<p>Nykypäivänä opetuksen tukena oppilaitoksissa hyödynnetään useita erilaisia verkkopalveluita. Yhteistä jokaiselle näistä palveluista on, että ne tallentavat tietoa käyttäjän toiminnasta sekä käyttäjän suorituksien arvioinneista. Tämän tutkielman tavoitteena on perehtyä mahdollisuuksiin hyödyntää tätä tietoa oppijan oppimistuloksien sekä opetuksen parantamiseksi.</p> <p>Tutkielmassa tarkastellaan oppimisen analysoimisen tarpeita ja miksi oppimisanalytiikkaa halutaan tehdä. Lisäksi tarkastellaan Moodlea datalähteenä oppimisanalytiikalle sekä tutustutaan tarkemmin kahteen erilaiseen oppimisanalytiikassa käytettyyn tietomalliin. Lopuksi tarkastellaan kuinka oppimisanalytiikan avulla voidaan tietolähteiden avulla muodostaa oppijaa sekä opetusta tukevia analyyskejä.</p>			
<p>ACM Computing Classification System (CCS) Information systems → Information systems applications → Decision support systems → Data analytics</p>			
Avainsanat — Nyckelord — Keywords			
learning analytics, Moodle			
Säilytyspaikka — Förvaringsställe — Where deposited			
Helsingin yliopiston kirjasto			
Muita tietoja — övriga uppgifter — Additional information			

HELSINGIN YLIOPISTO – HELSINGFORS UNIVERSITET – UNIVERSITY OF HELSINKI

Tiedekunta — Fakultet — Faculty		Koulutusohjelma — Utbildningsprogram — Study programme	
Faculty of Science		Bachelor's Programme in Computer Science	
Tekijä — Författare — Author			
Tuomas Alanen			
Työn nimi — Arbetets titel — Title			
Oppijan kehittymisen tukeminen oppimisanalytiikalla Moodlessa			
Ohjaajat — Handledare — Supervisors			
Lea Kutvonen			
Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages	
Bachelor's thesis	June 5, 2025	20 pages, 1 appendix pages	
Tiivistelmä — Referat — Abstract			
<p>Nowadays schools and universities are using many kinds of web services to support learning and teaching. These services have things in common: all of them are saving information about users' interactions and performance assessments. This thesis' plan is to dive into possibilities on how this data can be used to improve learners' outcome and improve teaching to be better.</p> <p>Thesis will take look into needs of learning analytics and the reasons why learning analytics is done. As an example, Moodle will be investigated as a data source and how two different kind of data model used in learning analytics are built. In the end there are some examples how learning analytics can be used with data sources to help learner and to make teaching supportive analytics.</p>			
<p>ACM Computing Classification System (CCS) Information systems → Information systems applications → Decision support systems → Data analytics</p>			
Avainsanat — Nyckelord — Keywords			
learning analytics, Moodle			
Säilytyspaikka — Förvaringsställe — Where deposited			
Helsinki University Library			
Muita tietoja — övriga uppgifter — Additional information			

Sisällys

1	Johdanto	1
2	Oppimisen analysoinnin tarpeet	3
3	Tietomallin muodostaminen oppijan kehittymisestä	5
3.1	Moodle datalähteenä	5
3.2	Yleistetty malli	7
3.3	Tietomallin rakentamisessa huomioitavaa	9
4	Tietomallin hyödyntäminen oppimisanalytiikassa	12
4.1	Yksittäiseen oppijaan kohdennetut ehdotukset	12
4.2	Opetuksen kehittämiseen kohdennetut ehdotukset	14
5	Yhteenveto	15
	Lähteet	17
A	Tekoälyn käyttö tutkielmassa	i

1 Johdanto

Nykyopetuksen tukena oppilaitoksissa hyödynnetään useita erilaisia verkkopalveluita. Tällaisia verkkopalveluita on esimerkiksi Moodle, ItsLearning tai Thinglink. Yhteistä näille palveluille on, että ne tallentavat tietoa käyttäjän toiminnasta sekä hänen tekemien suoritusten arvioinneista. Kuinka näiden järjestelmien keräämä tietoaaineisto voidaan hyödyntää oppijan avuksi? Tämän tutkielman tavoitteena on perehtyä mahdollisuuksiin hyödyntää tätä tietoa oppijan oppimistuloksien parantamiseksi.

Tämä tutkielma analysoi miten oppimisanalytiikkaa voidaan hyödyntää oppijan tukena. Erityisesti tarkastellaan mahdollisuuksia hyödyntää oppimisanalytiikkaa Moodle -verkkoppi-
misympäristössä. Oppimisanalytiikassa käytetyistä ennustavista malleista tarkastellaan Bayesin mallia ja linearista regressiota.

Tahtotilana on muodostaa käsitys siitä, kuinka oppijalle voidaan esittää hänen oma toimintansa oppimisen eteen selkeässä helposti ymmärrettävässä muodossa. Tämän tarkoituksena on tarjota mahdollisuus ymmärtää omaa toimintaa sekä antaa mahdollisuuksia kehittää omaa toimintaa sen seurauksena. Päästäksemme tähän halutaan löytää oppijan käyttäytymisestä erilaisia malleja, joiden avulla pystyttäisiin havaitsemaan esimerkiksi oppijan tarvitseman tuki tai löytämään esimerkiksi isommassa kuvassa tarve kohdistaa lisää opetusta osaamisvajeen paikkaamiseksi.

Tutkielman luku 2 tarkastelee oppimisen analysoimisen tarpeita ja miksi oppimisanalytiikkaa halutaan tehdä. Luvussa käsitellään oppimisanalytiikan rakennetta sekä kuinka oppimisanalytiikka toteutuu käytännössä. Muodostetaan käsitys kuinka oppimisanalytiikka muodostaa syklin, joka tuottaa koko ajan näkymän oppimiseen.

Luku 3 tarkastelee Moodlea datalähteenä oppimisanalytiikalle sekä tutustutaan tarkemmin kahteen erilaiseen oppimisanalytiikassa käytettyyn tietomalliin. Moodlen osalta nostetaan esille aktiviteettien merkitys datalähteenä analyysimalleille. Luvussa myös tarkastellaan Naiivin Bayesin mallin sekä regressioanalyysin toteuttamista.

Luku 4 tarkastelee kuinka oppimisanalytiikan avulla voidaan datalähteiden avulla muodostaa oppijaa sekä opetusta tukevia analyyskejä. Luvussa tuodaan esille kuinka oppija, opettaja sekä hallintohenkilöstö pystyy hyötymään oppimisanalytiikasta saatavasta analyysistä osana heidän tekemistä.

Tutkielma pyrkii vastaamaan tutkimuskysymyksiin *"Miten oppimisympäristöstä saatua tietoaaineistoa voidaan hyödyntää oppijan tukemiseksi oppimisanalytiikan avulla"* sekä *"Millaista tietoaaineistoa Moodlesta saadaan oppimisanalytiikan prosessin käyttöön?"* Ensimmäinen tutkimuskysymys käsittelee yleisesti oppimisanalytiikkaa ilmiönä ja kuinka sitä voidaan hyödyntää työkaluna. Toinen tutkimuskysymys pureutuu tarkemmin Moodleen lähdejärjestelmänä ja siihen, millaista dataa sieltä voitaisiin saada analysoitavaksi. Tutkimuskysymyksien taustalla on työkokemuksen tuoma ymmärrys Moodlen toiminnasta ja siitä miten järjestelmä on rakentunut oppijan sekä opettajan näkökulmasta.

Oppimisanalytiikka (Learning Analytics, LA) (Long ja Siemens, 2011) on tutkimusala, joka tutkii oppimista ja opettamista oppimisympäristöissä. Sen juuret ovat ihmisten ja koulutusjärjestelmän välisen vuorovaikutuksen analysoimisessa (Siemens, 2013). Oppimisanalytiikka jakautuu kahteen pääalueeseen menetelmiin ja sovelluksiin. Menetelmiin kuuluvat algoritmit ja tilastolliset mallit, joiden avulla analysoidaan kerättyä oppimisdataa. Sovelluksiin puolestaan kuuluu tavat, joilla menetelmien tuloksien avulla vaikutetaan ja kehitetään oppimistuloksia.

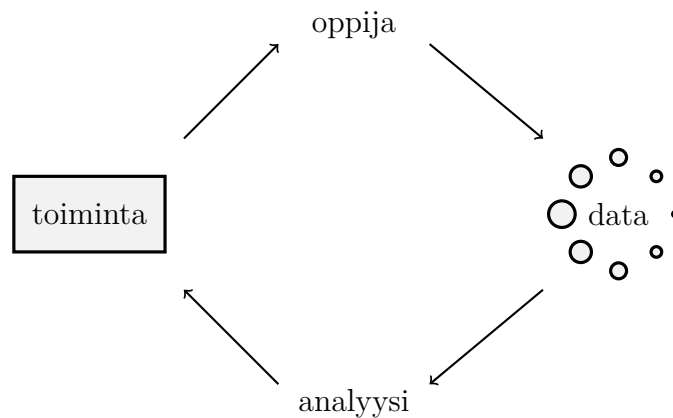
Oppimistiedon louhinta (Educational Data Mining, EDM) hyödyntää tilasto-, koneoppimis- ja datalouhintaalgoritmeja oppimisdatan käsittelyyn (Romero ja Ventura, 2010; Siemens, 2013). Oppimistiedon louhinta kehittää erityisesti oppimisdataalle soveltuvia menetelmiä ja tähtää erityisesti oppimisdataalle yksilöllisten datapisteiden hyödyntämiseen. Oppimisanalytiikka ja oppimistiedon louhinta ovat osittain päällekkäisiä tutkimusaloja etenkin menetelmien osalta. Kokonaiskuvassa oppimistiedon louhinta keskittyy enemmän menetelmien kehittämiseen, kun taas oppimisanalytiikka keskittyy analyttiseen lopputulokseen.

Verkko-oppimisympäristö (Learning Management System, LMS) on verkkopohjainen järjestelmä, joka mahdollistaa oppimateriaalin, opiskelijoiden toiminnan, tehtävätyökalujen ja oppijan edistymisen seurannan (Mohd Kasim ja Khalid, 2016) missä vain oppijan haluamalla päätelaitteella. Verkko-oppimisympäristöjen juuret on oppijoiden suorituksien taltioinnissa ja pitää yllä kurssille ilmoittautuneista oppijoista, mutta nykypäivänä järjestelmät mahdollistavat verkkopohjaisen oppimisen kokonaisuudessaan sekä pääsyn keskitetysti kurssin tietoihin. Verkko-oppimisympäristön käytön aikana muodostuu tietoa suorituksista ja käyttäytymisestä taustalla on tiedon tallentamiseen käytetty tietokanta, johon on ryhmitelty oppimisesta kerättyä tietoa (Romero et al., 2014).

2 Oppimisen analysoinnin tarpeet

Vakiintunut oppimisanalytiikan määritelmä on 1st International Conference on Learning Analyticsin määritelmä (Siemens, 2013; Clow, 2012): oppimisanalytiikka on oppijoista kerättävän datan mittaamista, keräämistä, analysointia ja raportointia, jota hyödynnetään oppimisen ja sen ympäristön ymmärtämiseen ja optimoimiseen.

Oppimisanalytiikkaa voidaan kuvata syklinä, jossa on neljä osa-aluetta: oppija, data, analyysi ja toiminta (Clow, 2012). On tilanteita, joissa syklistä jää toiminta pois. Tällaisia tilanteita ovat esimerkiksi raporttien muodostaminen oppimisdatasta, joiden pohjalta ei tehdä toimenpiteitä. Kuva havainnollistaa oppimisanalytiikan syklin osa-alueiden riippuvuuden toisistaan.



Kuva 2.1: Oppimisanalytiikan eri vaiheet tiivistetysti (Clow, 2012).

Syklissä oppija on oppimisanalytiikan lähtökohta (Clow, 2012). Oppijoista kerätään dataa analytiikkaa varten, jota koostetaan erilaisista lähteistä (Wolff et al., 2013). Oppimisympäristöstä saatavaa dataa voi olla esimerkiksi lokeihin kerätty tieto oppijan oppimisympäristössä liikkumisesta tai opintotietojärjestelmässä aiemmat kurssisuoritukset. Toisaalta myös oppijan oppimiskäyttäytymisen ja -tyylin ymmärtäminen ovat oppimisanalytiikassa keskiä (Hasan et al., 2020).

Oppijasta muodostunutta dataa voidaan tarkastella ja analysoida oppimisprosessin havainnollistamiseksi (Clow, 2012). Tämä on oppimisanalytiikan tärkein vaihe. Oppijoista saatavan datan perusteella voidaan tunnistaa esimerkiksi putoamisvaarassa olevia opiskelijoita tai ennustaa heidän menestymistä kurssilla.

Oppimisanalytiikassa syklin viimeisen kohdan, toiminnan, on tarkoitus vaikuttaa oppijaan (Clow, 2012). Toimintaa voi olla esimerkiksi oppijan käytössä oleva seurantanäkymä, jossa voi vertailla toisiin opiskelijoihin tai tarvittavan tuen kartoittaminen putoamisvaarassa olevalle opiskelijalle. Oppijan omien havaintojen ja opettajan havaintojen vaikutukset kohdistuvat oppijaan itseensä. Oppija voi hyödyntää analytiikasta saatavaa tietoa oman oppimisensa kehittämiseen hyvin nopeallakin vasteajalla.

Toiminta ei tavoita aina oppijaa, sillä tuloksia voidaan hyödyntää usealla tasolla (Clow, 2012). Opettaja voi hyödyntää aiemman kurssi-iteraation kurssiarvosanoja kurssin kehittämisen tukena. Kurssin aikana opettajat toimet voivat vaikuttaa yhden oppijan sijasta myös useampaan oppijaan, ja opettajan toiminnan vaikutukset eivät välttämättä ole heti havaittavissa.

Hallinto henkilöstön toiminnan vaikutukset ovat laajempia ja hitaammin havaittavia heidän yhdistäessä myös opettajalta saatavan palautteen analyysiinsa (Clow, 2012). Hallinto henkilöstö pystyy toiminnallaan vaikuttamaan isompaan joukkoon oppijoita kuin yksittäinen oppija esimerkiksi jakamalla kurssin kahteen osaan. Toisaalta oppimisanalytiikkaa voidaan hyödyntää laajemmalla tasolla esimerkiksi osana opetussuunnitelmatyötä, jolloin vaikutukset ovat vielä hitaammin havaittavissa, mutta niiden kattavuus on laajin (Clow, 2013).

Oppimisanalytiikkaa voidaan hyödyntää kolmessa eri käyttötarkoituksessa, joita ovat kuvaileva analytiikka, ohjaava analytiikka ja ennustava analytiikka (Auvinen, 2017; Daniel, 2015). Kuvailevassa analytiikassa kuvaillaan ja analysoidaan oppijoista sekä muista oppimisen osa-alueista saatavaa historiatietoa. Kuvaileva analytiikka etsii esimerkiksi nykyisiä oppimistrendejä. Ennustava analytiikka puolestaan tarjoaa oppilaitoksille mahdollisuuden tehdä datan perusteella parempia päätöksiä ja näkymiä nykytilasta. Tavoitteena on estimoida tulevien tapahtumien todennäköisyyksiä. Ohjaava analytiikka puolestaan tarjoaa oppilaitoksille mahdollisuuden arvioida nykyistä toimintaansa vaihtoehtoisten mallien pohjalta ja ohjaa parempiin päätöksiin.

3 Tietomallin muodostaminen oppijan kehittämisestä

Oppimisanalytiikassa yhdistelemällä tilastollisia menetelmiä ja ennustavaa mallintamista voidaan kohdentaa ohjausta oppijoiden haasteisiin oppimisessa ja tarjoamalla kohdistettua tukea saatavan datan avulla (Ranjeeth et al., 2020). Käytettävät ennustavat mallit voivat olla mitä vain datanlouhinta-, koneoppimis- ja keinoälymenetelmiä. Datalähteenä malleille voidaan hyödyntää eri oppijasta tietoa sisältäviä järjestelmiä, kuten verkko-oppimisympäristö Moodlea. Moodle tarjoaa esimerkiksi aktiviteeteistä laajasti erilaista tietoa hyödynnettäväksi mallin muodostamiseen.

3.1 Moodle datalähteenä

Moodle (Modular Object-Oriented Dynamic Learning System) on vuodesta 1999 lähtien kehitetty avoimen lähdekoodin verkko-oppimisympäristö, joka on julkaistu GPL-3.0 -lisenssillä (Dougiamas, 2021; Dougiamas, 2022). Moodlella on yli 315 miljoonaa käyttäjää eri puolilla maailmaa 178 tuhannella eri Moodle-sivustolla (Moodle.org, n.d.). Moodle on rakennettu käyttäen PHP-ohjelmointikieltä ja tiedon tallentamiseen relaatiotietokantaa. Suorat SQL-kyselyt tietokantaan ja Moodlen tarjoamat metodit mahdollistavat Moodlen keräämän tiedon hyödyntämisen osana data-analyysia. Moodlen tietokantarakenteesta löytyy selkeä indeksointi avaimien perusteella (Green, 2022), jonka perusteella tietokantataulusta toiseen asioiden jäljittäminen on mahdollista.

Moodlessa on vakiona 23 erilaista aktiviteettia, joista jokainen tallentaa erilaista tietoa tietokantaan (Dougiamas, 2022). Jokaisella aktiviteetillä on myös omia tietokantatauluja, joihin tallennetaan aktiviteettiin liittyvä tieto. Lisäksi Moodlen kehittäjäyhteisö on julkaissut paljon Moodlea laajentavia aktiveettejä (Moodle.org, 2022). Oppilaan osaamista mittaavia aktiviteettejä ovat esimerkiksi tentti, palaute, työpaja, oppitunti, keskustelualue ja H5P.

Esimerkiksi työpaja tallentaa kaikki suoritukset tauluun *workshop_submissions* ja suoritusten arvioinnit tauluun *workshop_grades* (Green, 2022). Työpaja mahdollistaa myös vertaisarvioinnin (taulussa *workshop_assessments*), jossa oppija joutuu arvioimaan omaa

ja toisten osaamista hyödyntämisen analytiikassa. Keskustelualueelta voidaan mitata oppijoiden aktiivisuutta viestien lukumäärällä (Mwalumbwe ja Mtebe, 2017). Aktiviteetistä myös saadaan tieto, onko sitä avattu kertaakaan taulun *course_module_completion* avulla.

Moodle tallentaa tietokantatauluun *logstore_standard_log*, eli tapahtumalokiin, kaikki Moodlen Event API:n kautta tulevat tapahtumat (Dougiamas, 2022; Dougiamas et al., 2021). Tapahtumien avulla voidaan kerätä tietoa toiminnasta verkko-oppimisympäristössä (Agudo-Peregrina et al., 2014). Lokitietoa erilaisista tapahtumista voi esimerkiksi tulla Moodlen ytimen komponenteista, eri aktiviteeteistä, työkaluista ja raporteista riippuen komponentin luonteesta. Useimmat aktiviteetit tallentavat lokiin merkittäviä tapahtumia, kuten suoritusten luomisen aktiviteettiin, kurssimoduulissa vierailun, tenttiin vastaamisen ja vertaisarvioinnin antamisen. Moodlen ytimessä oppijan kannalta tärkeimmät ovat kirjautumiseen ja kurssin katseluun liittyvät tapahtumat. Lokitietoihin tallentuu aina tieto kuka on vieraillut, milloin on vieraillut, missä on vieraillut ja mistä on vieraillut (Abdullah, 2015). Tapahtumalokin avulla voidaan tarkastella oppijoiden toiminnan painottumista eri kellonaikoihin.

Moodlen yhteisö on myös etsinyt erilaisia tapoja kerätä palautetta oppijoilta. Yksi tällainen on pikapalautetoiminnallisuus (*block_point_view*), joka antaa kolmiportaisen itsearviointimahdollisuuden aktiviteettikohtaisesti (Fombaron, 2021). Tämä mahdollistaa helpon ja nopean tavan saada oppijalta itsearviointidataa siitä, miten oppija itse näkee oman suoriutumisensa kyseisessä tehtävässä. Tietokantataulusta *block_point_view* pystytään hakemaan käyttäjän äänestystulos kurssin, kurssimoduulin tai käyttäjän perusteella.

Joidenkin tietojen, kuten oppijan tarkemman toiminnan seuraamiseen sivulla tarvitaan kolmannen osapuolen tuottamaa tekniikkaa (Filvå et al., 2014). Tällainen seuraamiseen soveltuva työkalu on esimerkiksi Google Analytics, joka seuraa tarkemmin käyttäjän toimintaa sivustolla. Moodlen lokitiedoista selviää milloin sivu on ladattu, mutta tietoa kuinka kauan oppija sivulla on todellisuudessa viettänyt aikaa ei tällä menetelmällä pystytäkään muodostamaan (Dougiamas, 2022). On teoreettisesti mahdollista, että oppija on avattu sivun katsellut sitä minuutin ajan ja tämän jälkeen lähtenyt kahville. Jos seuraava sivulataus on tunnin päästä, niin tästä ei pystytäkään luotettavasti laskemaan sivulla vietettyä todellista aikaa.

3.2 Yleistetty malli

Analytiikkamallin rakentaminen on iteratiivinen prosessi, jossa on useita vaiheita (Hämäläinen ja Vinni, 2010). Iteratiivisen prosessin aikana kokeillaan useita erilaisia malleja, datan esitysmuotoja ja algoritmien asetuksia parhaan mahdollisen tietomallin löytämiseksi. Valitun mallin toimivuus voidaan todentaa luokittelun onnistumisella, sillä sen soveltuvuus voidaan kyseenalaistaa liian monen luokitteluvirheen jälkeen.

Luokittelua eli asioiden lajittelemista eri kategorioihin hyödynnetään usein opetuksessa paljon, kuten opettajien arvioidessa oppijoiden tietotasoa, motivaatiota ja käytöstä (Hämäläinen ja Vinni, 2010). Yksi esimerkki luokittelusta on kokeiden ja tenttien arvostelu, missä vastauksia luokitellaan eri pistemäärin arviointiperusteiden mukaan. Oppimisanalytiikassa luokittelua tehdään selitettävän muuttujan arvoa ennustavalla mallilla, jota ennustetaan selittävien muuttujien arvojen avulla. Luokittimia voidaan tehdä joko käsityönä määritellen raja-arvoja tai yleisemmällä tavalla opettaa luokitin luokittelemaan olemassa olevan datan perusteella.

Useissa oppimisanalytiikkaa käsittelevissä tutkimuksissa on kokeiltu erilaisia luokittelualgoritmeja parhaiten toimivan mallin löytämiseksi (Akçapınar et al., 2019). Usein käytettyjä algoritmeja ovat naiivi Bayes, satunnaismetsä-malli, tukivektorikone (SVM), neuroverkko ja k-lähinaapurimenetelmä. Yksi tapa etsiä parhaiten toimivaa mallia on tehdä suorituskymmittauksia, joissa tarkastellaan tarkkuutta, herkkyyttä, yksityiskohtaisuutta ja F-mittaa.

Yksi tapa tehdä luokittelua on käyttää naiivia Bayesin luokitinta (Natingga, 2018). Naiivi Bayesin luokitin pohjautuu Bayesin teoreemaan

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)},$$

missä A ja B ovat tapahtumia, $P(A)$ on todennäköisyys tapahtumalle A olla tosi ja $P(A|B)$ on ehdollinen todennäköisyys tapahtumalle A olla tosi, mikäli tapahtuma B on tosi. Naiivissa Bayesin luokittimessa datapisteiden joukolle annetaan Bayesin teoreeman perusteella todennäköisin luokka. Tämä tapahtuu laskemalla todennäköisyys sille, kuinka todennäköisesti asia A tapahtuu, jos ehto B saa tietyn arvon.

Bayesin teoreemaa voidaan hyödyntää myös useamman todennäköisyystapahtuman kanssa, jolloin käytetään laajennettua Bayesin teoreemaa (Natingga, 2018). Jos määritellään tapahtumat B_1, \dots, B_n olemaan ehdollisesti riippumattomia tapahtumasta A , niin Baye-

sin teoreema voidaan esittää muodossa

$$P(A|B_1, \dots, B_n) = \frac{P(B_1, \dots, B_n|A) \cdot P(A)}{P(B_1, \dots, B_n)}.$$

Nämä satunnaismuuttujina toimivat todennäköisyystapahtumat voivat olla diskreettejä tai jatkuvia seuraten todennäköisyysjakaumaa, kuten normaalijakaumaa.

Käytettäessä Bayesilaista todennäköisyyttä täytyy vertailtavien tapahtumien olla riippumattomia toisistaan (Natingga, 2018). Jos vertaillaan lämpötilaa ja vuodenaikaa keskenään, niin näiden välillä havaitaan olevan riippuvuus: talvella on kylmää ja kesällä lämmin. Tämä estää Bayesin teoreeman käyttämisen luokitteluun. Tämä voidaan kiertää tekemällä analyysia niille data-aineiston tapahtumille, jotka eivät ole riippuvia toisistaan.

Toinen mahdollisuus tehdä tilastollista analyysia kerätylle oppimisdataalle on regressioanalyysi (Song, 2018; Romero ja Ventura, 2010; Papamitsiou ja Economides, 2014). Regressioanalyysiä voidaan tehdä usealla eri tavalla, kuten yksinkertaisella lineaarisella regressiolla, usean selittäjän lineaarisella regressiolla ja logistisella regressiolla. Regression avulla voidaan ennustaa lineaarisesti esimerkiksi kuinka opiskelija tulee menestymään eri selittävien muuttujien vaikutus huomioiden.

Lineaarinen regressio kuvaa yhden selittävän ja yhden selitettävän muuttujan yhteyttä toisiinsa (Ross, 2017). Yksinkertainen lineaarinen regressio voidaan esittää kaavana

$$Y = \alpha + \beta x + e,$$

jossa x kuvaa selittävää muuttujaa ja y kuvaa selitettävää muuttujaa. Parametrit α ja β ovat tuntemattomia suureita, estimaattoreita, jotka estimoidaan datan perusteella. Muuttuja e kuvaa satunnaista virhettä, jonka oletetaan noudattavan normaalijakaumaa odotusarvolla 0 ja varianssilla σ^2 . Varianssin oletetaan olevan sama riippumatta selittävistä muuttujista x .

Parametrien α ja β estimointiin voidaan käyttää pienimmän neliösumman estimointia (Ross, 2017). Tällöin halutaan löytää sellaiset arvot estimaateille α ja β , joilla virheen neliösumma $\sum_{i=1}^n \epsilon_i^2$ on mahdollisimman pieni. Pienimmän neliösumman estimaatit $\hat{\alpha}$ ja $\hat{\beta}$ parametreille α ja β saadaan laskettua kaavoista

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

ja

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x},$$

missä $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ ja $\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$.

Estimoidussa regressioviivassa $y = \hat{\alpha} + \hat{\beta}x$ estimaatti $\hat{\alpha}$ kuvaa suoran kulmakerrointa ja estimaatti $\hat{\beta}$ suoran vakiota, eli kohtaa y-akselilta missä suora leikkaa y-akselin (Ross, 2017). Tämän estimoidun regressioviivan avulla voidaan ennustaa selitettävän muuttujan y arvoja käyttäen selittävän muuttujan x arvoja.

Yksinkertainen lineaarinen regressio voidaan laajentaa usean selittäjän lineaariseksi regressioksi, joka kuvaa useamman selittävän muuttujan x_1, \dots, x_i vaikutusta selitettävään muuttuajaan Y (Ross, 2017). Matemaattisena kaavana esitettynä usean selittäjän lineaarinen regressio on

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + e,$$

jossa Y on selitettävä muuttuja, ja x_i kuvaa selittäviä muuttujia, missä $i = 1, \dots, k$. Regressioparametrejä yhtälössä kuvaa $\beta_0, \beta_1, \dots, \beta_k$ ja satunnaisvirhettä e .

Myös regressiossa täytyy selittävien muuttujien olla riippumattomia toisistaan, eli nämä muuttujat eivät saa olla keskenään korreloivia (Daoud, 2017). Tätä ilmiötä kutsutaan multikollineaarisuudeksi. Ilmiö voidaan havaita tapauksissa, joissa tapahtuu suurta vaihtelua estimoiduissa kertoimissa lisättäessä tai poistettaessa selittäviä muuttujia tai poistettaessa yksittäisiä datapisteitä.

3.3 Tietomallin rakentamisessa huomioitavaa

Ennen datan syöttämistä analysointia tai luokittelua tekeväälle mallille, tulee aineistolle suorittaa esikäsittely (Romero et al., 2014). Esikäsittely aloitetaan keräämällä tarvittava data, joka ryhmitellään sopiviin ja järkeviin kokonaisuuksiin. Datan ryhmittelyn jälkeen poistetaan siitä kaikki epäolennainen ja virheellinen sisältö. Aineistosta tunnistetaan käyttäjät ja heidän asiointisessiot kohdistuaksemme analyysin oikeisiin oppijoihin. Korreloivat ja toisteiset muuttujat jätetään pois, kun valitaan aineistosta sopivat selittävät muuttujat. Isoista data-aineistoista poistetaan aiempien vaiheiden jälkeen turhiksi jääneet kentät, jotka olisivat epäolennaisia prosessille. Lopuksi tarkastellaan mahdollisuutta muodostaa uusia muuttujia olemassa olevien muuttujien perusteella, kuten normalisoida jonkin muuttujan arvot tietyllä välillä tai muuttaa esitystapaa sopivammaksi.

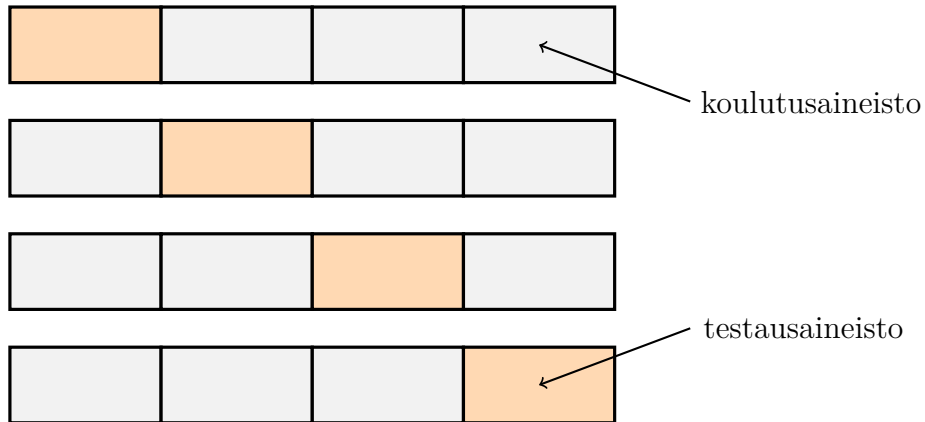
Kun yhdistellään tietyllä välillä liikkuvia muuttujia kategoristen muuttujien kanssa, voidaan käyttää tekumuuttujia kuvaamaan näitä arvoja (Ross, 2017). Tällöin voidaan hyödyntämään sellaisia selittäviä kategorisia muuttujia, jotka eivät lähtökohtaisesti ole nu-

meerisessä muodossa. Jos esimerkiksi usean selittäjän lineaarisessa regressiossa muuttuja x_3 kuvaa onko oppija tutkinto-opiskelija, voidaan tämä esittää numeraalisessa muodossa seuraavasti:

$$x_3 = \begin{cases} 1 = \text{oppija on tutkinto-opiskelija} \\ 0 = \text{oppija ei ole tutkinto-opiskelija} \end{cases}.$$

Mallia toteutettaessa on huomioitava yli- ja alisovittamisen vaara, jotta mallin tarkkuus ei kärsisi (Hämäläinen ja Vinni, 2010). Ylisovittamisessa malli on sovitettu koulutusaineistoon niin tarkasti, että se huomioi jopa kaikki erikoistapaukset sekä koulutusdatan virheet. Tämä ilmenee liian monimutkaisena mallina suhteessa käytettävän data-aineiston kokoon. Alisovittamisessa liian yksinkertainen malli ei pysty välttämättä tulkitsemaan data-aineistoa ja täten malli ei kuvaa todellisuutta tai kuvaa sitä todella vähän.

Yksi tapa jakaa data-aineisto koulutus- ja testidataan on käyttää ristivalidointia, kuten k -kertaista ristiinvalidointia (Deisenroth et al., 2020). Aineisto jaetaan k osaan, joista yhtä osaa kerrallaan käytetään testiaineistona \mathcal{V} ja $k - 1$ osaa koulutusaineistona \mathcal{R} . Tällöin aineistosta käytetään suurin osa mallin kouluttamiseen, mutta samasta aineistosta saadaan myös testiaineisto muodostettua. Ristiinvalidoinnissa käydään läpi kaikki mahdolliset k vaihtoehtoa valita testiaineisto jakamalla data-aineisto kahteen osaan $D = \mathcal{R} \cup \mathcal{V}$, missä $\mathcal{R} \cap \mathcal{V} = \emptyset$. Näiden k -suorituskerran muodostamien mallien suorituskkyä tarkastellaan keskiarvona.



Kuva 3.1: Ristiinvalidoinnissa data-aineisto jaetaan kerrallaan k osaan, missä $k - 1$ osaa ovat koulutusaineistoa (harmaalla merkityt osuudet) ja yksi osa testausaineistoa (oranssilla merkitty osuus) (Deisenroth et al., 2020).

Koulutusaineistolla \mathcal{R} koulutetun mallin f suorituskkyä tarkastellaan testausaineiston \mathcal{V} avulla, jolle lasketaan keskineliövirheen neliöjuuren avulla empiirinen riski testausaineistolla \mathcal{V} (Deisenroth et al., 2020). K -kertaisessa ristiinvalidoinnissa lasketaan jokaiselle

koulutusaineiston k -osan $\mathcal{R}^{(k)}$ predikaattorille $f^{(k)}$ empiirinen riski $R(f^{(k)}, \mathcal{V}^{(k)})$ käyttäen testiaineistoa $\mathcal{V}^{(k)}$. Kaikille mahdollisille k -osaan jaoille ristiinvaldointi arvioi odotetun yleistysvirheen kaavasta

$$\mathbb{E}_{\mathcal{V}}[R(f, \mathcal{V})] \approx \frac{1}{K} \sum_{k=1}^K R(f^{(k)}, \mathcal{V}^{(k)}).$$

Käytettävässä arvioinnissa on kaksi lähdettä, joista toisessa rajatulla koulutusaineistolla ei välttämättä saada parasta mahdollista $f^{(k)}$ ja toisessa testausaineistolla ei saada tarkkaa arviota riskistä $R(f^{(k)}, \mathcal{V}^{(k)})$.

4 Tietomallin hyödyntäminen oppimisanalytiikassa

Oppimisanalytiikan hyödyntäjiä on useissa käyttäjäryhmissä: oppijat, opettajat, opetuksen tutkijat, organisaatiot ja hallintohenkilöstö (Romero ja Ventura, 2010). Nämä käyttäjäryhmät hyödyntävät samaa tietoaaineistoa eri käyttökohteisiin jalostettuna, joita on esimerkiksi oppijan osalta oman oppimisen parantaminen tai opettajien osalta opetuksen parantaminen vastaamaan paremmin oppijoiden vaatimuksia. Opetuksen parantamisessa tietoaaineiston avulla voidaan esimerkiksi parantaa opetukseen liittyvää päätöksentekoa löytämällä kurssit, joista on eniten hyötyä oppijoiden kehittymisen kannalta. Oppija puolestaan voi esimerkiksi saada vinkkejä hänen oppimispolkuaan hyödyntävistä oppimateriaaleista.

Ennustavien mallien avulla muodostetaan keskimääräistä oppijaa kuvaavia malleja, joiden avulla yksittäisiä oppijoita voidaan vertailla (Wolff et al., 2013). Ennustavan mallinnuksen avulla voidaan ennustaa esimerkiksi kuinka oppija tulee menestymään kurssilla ja onko oppija pääsemässä kurssia läpi. Tämä tapahtuu vertailemalla oppijaa muodostettuun malliin ja mallin avulla saatavaan ennusteeseen onko oppija vaarassa olla läpäisemättä kurssia. Tämän avulla oppijalle voidaan kohdistaa tarvittavia tukitoimia.

4.1 Yksittäiseen oppijaan kohdennetut ehdotukset

Oppijan menestymistä kurssilla voidaan ennustaa eri tarkoituksiin (Barber ja Sharkey, 2012a). Tähän voidaan hyödyntää aiempien kurssien menestystietoa muista lähteistä, sekä kurssin edistyessä lisätä kurssisuorituksista saatavaa tietoa mukaan analyysiin. Yhdistämällä tämän visualisointiin, voidaan tarjota oppijalle reaaliaikainen näkymä kurssimenestyksestä ja hyödyntää tätä motivaation lähteenä. Kurssitasolla voidaan seurata opiskelijan toimintaa kurssilla ja tehdä havaintoja kurssin edistymisestä ja menestymisestä (Long ja Siemens, 2011; Siemens, 2013). Tätä voidaan tehdä esimerkiksi luokittimilla tai yleistetyillä malleilla.

Naiivilla Bayesin luokittimella voidaan tunnistaa opiskelijoita, jotka ovat vaarassa saada hylätyn osallistumaltaan kurssilta. Esimerkiksi Barber ja Sharkey (2012b) käyttävät se-

littävinä muuttujina henkilöön liittyviä taustatietoja, kuten suoritettujen opintopisteiden suhdetta yritettyihin opintopisteisiin ja toimintaa verkko-oppimisympäristön keskustelueueella. Riippuen kurssin viikosta selittävälle muuttujille käytetään eri painoarvoja.

Kurssilta putoamisvaaraa voidaan ennustaa oppijan menestyksen perusteella tarkastelemalla riskitekijöitä oppijan toiminnasta verkko-oppimisympäristössä (Olivé et al., 2018; Suhonen ja Kinnari-Korpela, 2019). Tarkastelua voidaan laajentaa eri kurssien väliseksi (Kinnari-Korpela ja Suhonen, 2020) ja analytiikan löytäessä putoamisvaarassa olevan oppijan riskitekijöistä, voidaan hänelle tarjota kohdistetusti tukea oppimiseen jo aikaisessa vaiheessa.

Usean selittäjän lineaarista regressiota voidaan hyödyntää esimerkiksi etsittäessä eri relaatioita oppijoiden verkko-oppimisympäristön toiminnan ja akateemisen menestyksen väliltä (Agudo-Peregrina et al., 2014). Riippumattomia selittäviä muuttujia ovat eri tyyppiset interaktiot verkko-oppimisympäristössä ja riippuvana selittävänä muuttujana jokaisen oppijan saamana kurssin päättöarvosanana esittety akateeminen menestys. Näiden väliltä löydettiin merkittäviä relaatioita.

Oppimisanalytiikkaa voi hyödyntää myös sisällön suosittelemiseen (Long ja Siemens, 2011; Siemens, 2013). Tässä oppijan oppimispolku muotoillaan osaamista vastaavaksi esimerkiksi ohjaamalla perusasiat jo hyvin osaava oppija haasteellisemmalle kurssille tai tarjotaan heikommin pärjäävälle opiskelijalle taitotasoa vastaavia tehtäviä. Tämä voidaan jatkojaloistaa useammalle eri tasolle: mukautetaan oppimista oppijoiden mukaan prosessinäkö-kulmasta. Tätä voidaan jatkaa yksittäisen kurssin sisällön mukauttamisella ja jopa älykkään mukautuvan opintosuunnitelman luomiseen jatkumona kurssilta toiselle esimerkiksi oppimispolun muodossa.

Oppimispolku kertoo oppimistehtävien ohjeistukset ja tavoitteet, sekä havainnollistaa oppimisen edistymistä kurssin aikana (Toivola, 2017). Oppimispolun halutaan mahdollistaa oppijan oman luontaisen oppimistahdin hyödyntäminen. Oppimisanalytiikan avulla voidaan visualisoida oppijan edistyminen oppimispolulla ja tarjota myös suosituksia seuraavista tehtävistä (Long ja Siemens, 2011). Jos oppija ei ole vielä ymmärtänyt jotain oppimispolun osa-aluetta, voi analytiikka ehdottaa lisätehtävää osaamisen vahvistamiseksi ennen seuraavaan osa-alueeseen siirtymistä. Oppijan omaa käsitystä osaamisesta voidaan hyödyntää analytiikan tukena itsearviointin avulla. Oppijalle voidaan tarjota mahdollisuus vertailla omaa suorittamista anonymisoituun yhteenvetoon vertaisista oppijoista.

4.2 Opetuksen kehittämiseen kohdennetut ehdotukset

Oppimisanalytiikasta saatavalla datalla voidaan kehittää resurssien sijoittelua ja käyttöä (Long ja Siemens, 2011). Oppimisanalytiikalla voidaan löytää nykyisistä kursseista heikkoja kohtia, joihin ratkaisu voi olla uuden kurssin luominen tai nykyisen kehittäminen tukemaan osaamisvajeen paikkaamista. Tämä voi näkyä oppimateriaalin kehittämisenä, mikäli analytiikka osoittaa tietyn osa-alueen tehtävistä suoriutumisen olevan muita heikompi. Vertailussa voidaan hyödyntää eri opetusmateriaalin osien tarkastelukertoja.

Oppimisanalytiikan avulla voidaan ymmärtää paremmin oppilaitoksen onnistumisia ja haasteita (Long ja Siemens, 2011). Oppilaitoksen tuottavuutta ja tehokkuutta voidaan kehittää hyödyntämällä viimeisintä tietoa ja haasteisiin pystytään vastamaan nopeasti. Tätä voidaan toteuttaa analysoimalla erilaisia skenaariovaihtoehtoja kokeilemalla vaihtoehtoja oppijan opetukseen vaikuttavista tekijöistä, ja saada näiden pohjalta muodostetusta ennusteesta vertailtavaksi eri muutosten vaikutukset.

Oppimisanalytiikan avulla voidaan kehittää opetussuunnitelmia sekä tukea hallinnon päätöksiä (Romero ja Ventura, 2010; Long ja Siemens, 2011). Kurssivalikoimaa voidaan parantaa oppijoiden valintojen sekä suorituskertojen perusteella. Lisäksi voidaan löytää ajankohdaisia tarvepohjaisia lähtöjä uusille kursseille opetussuunnitelmaan, joille on tunnistettu tietoaineiston perusteella oppijalähtöinen tarve. Hallinnon päätöksissä voidaan parantaa esimerkiksi opetusresurssien kohdistamista niille kursseille, joille on tunnistettu tarvetta.

Yksi ennustamisen mahdollisuus on tarkastella valmistuuko koulutukseen hakija ennusteen mukaan tavoiteaikataulussa (Barber ja Sharkey, 2012a). Tämän toteuttamisessa hyödynnetään hiljaisia signaaleja, joita pystytään tunnistamaan jo ennen oppijan epäonnistumista tai katoamista kurssilta. Näitä hiljaisia signaaleja on esimerkiksi poissaolot luennoilta tai verkkokurssien keskustelualueita, huolimattomasti tehdyt tehtävät ja muutokset oppijan käytöksessä.

5 Yhteenveto

Oppimisanalytiikkaa voidaan kuvata oppijoista kerättävän tietoaaineiston mittaamisena, keräämisenä, analysointina ja raportointina (Siemens, 2013; Clow, 2012). Tätä hyödynnetään oppimisen ja sen ympäristön ymmärtämiseen sekä optimoimiseen.

Oppimisanalytiikkaa voidaan yleensä kuvata neljän osa-alueen syklinä: oppija, data, analyysi ja toiminta (Clow, 2012). Oppimisanalytiikan lähtökohta on oppija, joka tuottaa oppimisanalytiikassa käytettävän datan (Wolff et al., 2013). Kerättävä data voi olla esimerkiksi oppimisympäristöistä. Kerätyn datan pohjalta voidaan tarkastella ja analysoida oppimisprosessin havainnollistamiseksi, joka on oppimisanalytiikan tärkein vaihe. Toiminnan avulla on tarkoitus vaikuttaa oppijaan ja tarjota mahdollisuuksia kehittää omaa oppimistaan.

Oppimisanalytiikassa tuloksien hyödyntäjiä voi olla useita, eikä toiminta aina kohdistu oppijaan (Clow, 2012). Oppimisanalytiikasta saatavia tietoja voidaan hyödyntää oppijan toiminnan kehittämisen lisäksi opettajan toiminnan kehittämiseen, hallintohenkilöstötasolla esimerkiksi kohdistettaessa resursseja tai vielä laajemmalla tasolla esimerkiksi osana opetussuunnitelmatyötä (Clow, 2013). Tällöin voidaan havaita, että oppimisanalytiikka voi vaikuttaa sekä yhden tietyn oppijan toimintaan, mutta laajemmin myös useiden oppijoiden toiminnan kehittämiseen. Kehittämällä raportointia kohti enemmän datalla johtamista, pystytään tukemaan oppijoita oppimisanalytiikalla saavuttamaan tavoitteitaan paremmin.

Tilastollisten mallien sekä ennustavan mallintamisen avulla voidaan oppimisanalytiikan avulla tarjota ohjausta oppijoiden oppimishaasteisiin sekä tarjota kohdistettua tukea tietoaaineiston avulla (Ranjeeth et al., 2020). Eri oppijan tietoa sisältävistä järjestelmistä, kuten Moodlesta saatavaa tietoaaineistoa voidaan käsitellä yleistettyjen mallien, kuten luokittimien ja regressioanalyysin, avulla.

Yksi esimerkki oppimisanalytiikan tietoaaineiston lähteeksi on Moodle. Moodle on avoimen lähdekoodin verkko-oppimisympäristö, jolla on yli 315 miljoonaa käyttäjää eri puolilla maailmaa (Dougiamas, 2021; Dougiamas, 2022; Moodle.org, n.d.). Moodle tarjoaa useita erilaisia aktiviteetteja, jotka tallentavat tietoa oppijan toiminnasta relaatiotietokantaan. Oppimisanalytiikan osalta tentti, palaute, työpaja, oppitunti, keskustelualue ja H5P tuottavat konkreettista tietoa oppijan suoritumisesta. Moodlen tapahtumalokiin tallennetut

tapahtumat tuottavat oppijan toiminnan kannalta merkittävää tietoaaineistoa esimerkiksi tehtävien avauskertojen sekä oppijan toiminnan ajoittumisen osalta (Dougiamas et al., 2021; Abdullah, 2015).

Järjestelmistä saatavaa tietoaaineistoa voidaan käsitellä ennustaviin malleihin pohjautuvilla luokittimilla (Hämäläinen ja Vinni, 2010). Bayesin teoreemalla voidaan laskea todennäköisyystapahtumien avulla kunika todennäköisesti jokin asia A tapahtuu, mikäli ehto B saa tietyn arvon (Natingga, 2018). Toinen mahdollinen ennustava malli on käyttää regressioanalyysiä, kuten lineaarista regressiota (Ross, 2017). Se kuvaa yhden selitettävän ja yhden tai useamman selittävän muuttujan välistä yhteyttä toisiinsa.

Valmisteltaessa tietomallia on huomioitava myös aineiston käsitteleminen ennen sen syöttämistä ennustavalle mallille (Romero et al., 2014; Ross, 2017). Esikäsittelyssä kerätään kaikki tarvittava tietoaaineisto ja tämän jälkeen tietoaaineistoa ryhmitellään, siistitään ja muokataan analyysiin sopivaan muotoon. Lisäksi tarkastellaan mahdollisten tekemuuttujien tarve.

Oppimisanalytiikan avulla voidaan kohdentaa ehdotuksia yksittäisiin oppijoihin esimerkiksi tunnistamalla mahdollisesti kurssin reuttavia oppijoita (Barber ja Sharkey, 2012b) sekä etsimällä relaatioita oppijoiden verkko-oppimisympäristön toiminnan ja akateemisen menestyksen väliltä (Agudo-Peregrina et al., 2014). Koko koulupolkua voidaan tarkastella ennustamalla valmistuuko koulutukseen hakija tavoiteaikataulussa (Barber ja Sharkey, 2012a).

Opetuksen kehittämiseen voidaan oppimisanalytiikkaa soveltaa esimerkiksi kehittämällä resurssien sijoittelua ja käyttöä (Long ja Siemens, 2011; Romero ja Ventura, 2010). Käytännössä voidaan hyödyntää esimerkiksi kurssien heikkojen kohtien, oppilaitoksen onnistumisien ja haasteiden löytämiseen. Lisäksi voidaan hyödyntää opetussuunnitelmien sekä hallinnon päätöksiin tukemiseen. Näiden avulla voidaan kohdistaa resurssit vastaamaan todellista tarvetta sekä tarkastella onko jollekin osa-alueelle kohdistettava enemmän opetusresursseja.

Oppimisanalytiikalla ei voida korvata oppijoiden ohjausta, vaan oppimisanalytiikka on yksi työkalu kaikkien muiden työkalujen joukossa (Auvinen, 2017). Oppimisanalytiikalla voidaan tehostaa oppimisprosessia. Se tulee ymmärtää työvälineenä, jonka avulla ymmärretään paremmin oppimista ilmiönä, koska se havainnollistaa mitä oppimisessa tapahtuu. Oppijan on helpompi ymmärtää oppimisessa olevia haasteita, kun ne pystytään esittämään konkreettisemmin.

Lähteet

- Abdullah, M. A. (maaliskuu 2015). "Learning Style Classification Based on Student's Behavior in Moodle Learning Management System". *Transactions on Machine Learning and Artificial Intelligence* 3.1, s. 28–28. ISSN: 2054-7390. DOI: [10.14738/tmlai.31.868](https://doi.org/10.14738/tmlai.31.868).
- Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M. Á. ja Hernández-García, Á. (helmikuu 2014). "Can We Predict Success from Log Data in VLEs? Classification of Interactions for Learning Analytics and Their Relation with Performance in VLE-supported F2F and Online Learning". *Computers in Human Behavior* 31, s. 542–550. ISSN: 0747-5632. DOI: [10.1016/j.chb.2013.05.031](https://doi.org/10.1016/j.chb.2013.05.031).
- Akçapınar, G., Altun, A. ja Aşkar, P. (lokakuu 2019). "Using Learning Analytics to Develop Early-Warning System for at-Risk Students". *International Journal of Educational Technology in Higher Education* 16.1, s. 40. ISSN: 2365-9440. DOI: [10.1186/s41239-019-0172-z](https://doi.org/10.1186/s41239-019-0172-z).
- Auvinen, A.-M. (elokuu 2017). "Oppimisanalytiikka tulee – oletko valmis?" *Suomen eOppimiskeskus ry*. URL: <https://poluttamo.fi/2017/08/02/oppimisanalytiikka-tulee-oletko-valmis/>.
- Barber, R. ja Sharkey, M. (huhtikuu 2012a). "Course Correction: Using Analytics to Predict Course Success". *ACM International Conference Proceeding Series*. DOI: [10.1145/2330601.2330664](https://doi.org/10.1145/2330601.2330664).
- Barber, R. ja Sharkey, M. (huhtikuu 2012b). "Course Correction: Using Analytics to Predict Course Success". *ACM International Conference Proceeding Series*. DOI: [10.1145/2330601.2330664](https://doi.org/10.1145/2330601.2330664).
- Clow, D. (huhtikuu 2012). "The Learning Analytics Cycle: Closing the Loop Effectively". Teoksessa: *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. Vancouver British Columbia Canada: ACM, s. 134–138. ISBN: 978-1-4503-1111-3. DOI: [10.1145/2330601.2330636](https://doi.org/10.1145/2330601.2330636).
- Clow, D. (elokuu 2013). "An overview of learning analytics". *Teaching in Higher Education* 18.6, s. 683–695. ISSN: 1470-1294. DOI: [10.1080/13562517.2013.827653](https://doi.org/10.1080/13562517.2013.827653).
- Daniel, B. (2015). "Big Data and Analytics in Higher Education: Opportunities and Challenges". *British Journal of Educational Technology* 46.5, s. 904–920. ISSN: 1467-8535. DOI: [10.1111/bjet.12230](https://doi.org/10.1111/bjet.12230).

- Daoud, J. I. (joulukuu 2017). "Multicollinearity and Regression Analysis". *J. Phys.: Conf. Ser.* 949, s. 012009. ISSN: 1742-6588, 1742-6596. DOI: [10.1088/1742-6596/949/1/012009](https://doi.org/10.1088/1742-6596/949/1/012009).
- Deisenroth, M. P., Faisal, A. A. ja Ong, C. S. (2020). *Mathematics for Machine Learning*. Cambridge University Press.
- Dougiamas, M. (2021). "The Power of Open Educational Technology". Teoksessa: *Reimagining Digital Learning for Sustainable Development*. Routledge. ISBN: 978-1-00-308969-8.
- Dougiamas, M. (toukokuu 2022). *Moodle*. URL: <https://github.com/moodle/moodle> (viitattu 11.05.2022).
- Dougiamas, M., Mudrak, D., Lafuente, E., Monllaó, D., Bannister, T., de Raadt, M., Morris, R., Agarwal, A., Kavalerchik, N., Glancy, M., Škoda, P., Taneja, R., Massart, F., Greeve, A., Nelson 2, M., Guillaume, A. ja Massart, F. (heinäkuu 2021). "Logging 2 - MoodleDocs". *Logging 2*. URL: https://docs.moodle.org/dev/Logging_2 (viitattu 11.05.2022).
- Filvà, D. A., Guerrero, M. J. C. ja Forment, M. A. (kesäkuu 2014). "Google Analytics for Time Behavior Measurement in Moodle". Teoksessa: *2014 9th Iberian Conference on Information Systems and Technologies (CISTI)*, s. 1–6. DOI: [10.1109/CISTI.2014.6877095](https://doi.org/10.1109/CISTI.2014.6877095).
- Fombaron, Q. (marraskuu 2021). *Moodle Plugin « Point of View »*. URL: https://github.com/QuentinFombaron/moodle-block_point_view (viitattu 11.05.2022).
- Green, M. (maaliskuu 2022). *Moodle_3.11 Database*. URL: <https://www.examulator.com/er/output/index.html> (viitattu 11.05.2022).
- Hasan, R., Palaniappan, S., Mahmood, S., Abbas, A., Sarker, K. U. ja Sattar, M. U. (tammikuu 2020). "Predicting Student Performance in Higher Educational Institutions Using Video Learning Analytics and Data Mining Techniques". *Applied Sciences* 10.11, s. 3894. ISSN: 2076-3417. DOI: [10.3390/app10113894](https://doi.org/10.3390/app10113894).
- Hämäläinen, W. ja Vinni, M. (lokakuu 2010). "Classifiers for Educational Data Mining". Teoksessa: *Handbook of Educational Data Mining*. Toim. C. Romero, S. Ventura, M. Pechenizkiy ja R. Baker. Vol. 20103384. CRC Press, s. 57–74. ISBN: 978-1-4398-0457-5 978-1-4398-0458-2. DOI: [10.1201/b10274-7](https://doi.org/10.1201/b10274-7).
- Kinnari-Korpela, H. ja Suhonen, S. (heinäkuu 2020). *Oppimisanalytiikalla tehokkaampaan ohjaukseen*. URL: <https://sites.tuni.fi/tamk-julkaisut/tekniikka/oppimisanalytiikalla-tehokkaampaan-ohjaukseen/> (viitattu 11.05.2022).

- Long, P. ja Siemens, G. (syyskuu 2011). "Penetrating the Fog: Analytics in Learning and Education". *EDUCAUSE Review*, s. 30–32, 34, 36, 38, 40. URL: <https://er.educause.edu/-/media/files/article-downloads/erm1151.pdf>.
- Mohd Kasim, N. N. ja Khalid, F. (kesäkuu 2016). "Choosing the Right Learning Management System (LMS) for the Higher Education Institution Context: A Systematic Review". *International Journal of Emerging Technologies in Learning (iJET)* 11.06, s. 55. ISSN: 1863-0383. DOI: [10.3991/ijet.v11i06.5644](https://doi.org/10.3991/ijet.v11i06.5644).
- Moodle.org (marraskuu 2022). *Moodle Plugins Directory - Activity Modules*. URL: <https://moodle.org/plugins/?q=type:mod> (viitattu 11. 05. 2022).
- Moodle.org (n.d.). *Moodle Statistics*. URL: <https://stats.moodle.org/> (viitattu 11. 05. 2022).
- Mwalumbwe, I. ja Mtebe, J. S. (2017). "Using Learning Analytics to Predict Students' Performance in Moodle Learning Management System: A Case of Mbeya University of Science and Technology". *THE ELECTRONIC JOURNAL OF INFORMATION SYSTEMS IN DEVELOPING COUNTRIES* 79.1, s. 1–13. ISSN: 1681-4835. DOI: [10.1002/j.1681-4835.2017.tb00577.x](https://doi.org/10.1002/j.1681-4835.2017.tb00577.x).
- Natingga, D. (2018). *Data Science Algorithms in a Week - Second Edition*. 2nd edition. Packt Publishing. ISBN: 1-78980-607-0.
- Olivé, D. M., Huynh, D. Q., Reynolds, M., Dougiamas, M. ja Wiese, D. (lokakuu 2018). "A Supervised Learning Framework for Learning Management Systems". Teoksessa: *Proceedings of the First International Conference on Data Science, E-learning and Information Systems*. DATA '18. New York, NY, USA: Association for Computing Machinery, s. 1–8. ISBN: 978-1-4503-6536-9. DOI: [10.1145/3279996.3280014](https://doi.org/10.1145/3279996.3280014).
- Papamitsiou, Z. ja Economides, A. A. (2014). "Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence". *Journal of Educational Technology & Society* 17.4, s. 49–64. ISSN: 1176-3647.
- Ranjeeth, S., Latchoumi, T. P. ja Paul, P. V. (tammikuu 2020). "A Survey on Predictive Models of Learning Analytics". *Procedia Computer Science*. International Conference on Computational Intelligence and Data Science 167, s. 37–46. ISSN: 1877-0509. DOI: [10.1016/j.procs.2020.03.180](https://doi.org/10.1016/j.procs.2020.03.180).
- Romero, C., Romero, J. R. ja Ventura, S. (2014). "A Survey on Pre-Processing Educational Data". Teoksessa: *Educational Data Mining*. Toim. A. Peña-Ayala. Vol. 524. Cham: Springer International Publishing, s. 29–64. ISBN: 978-3-319-02737-1 978-3-319-02738-8. DOI: [10.1007/978-3-319-02738-8_2](https://doi.org/10.1007/978-3-319-02738-8_2).
- Romero, C. ja Ventura, S. (marraskuu 2010). "Educational Data Mining: A Review of the State of the Art". *IEEE Transactions on Systems, Man, and Cybernetics, Part C*

- (*Applications and Reviews*) 40.6, s. 601–618. ISSN: 1558-2442. DOI: [10.1109/TSMCC.2010.2053532](https://doi.org/10.1109/TSMCC.2010.2053532).
- Ross, S. M. (tammikuu 2017). ”Introductory Statistics”. Teoksessa: *Introductory Statistics (Fourth Edition)*. Toim. S. M. Ross. Oxford: Academic Press, s. 797–800. ISBN: 978-0-12-804317-2. DOI: [10.1016/B978-0-12-804317-2.00031-X](https://doi.org/10.1016/B978-0-12-804317-2.00031-X).
- Siemens, G. (lokakuu 2013). ”Learning Analytics: The Emergence of a Discipline”. *American Behavioral Scientist* 57.10, s. 1380–1400. ISSN: 0002-7642. DOI: [10.1177/0002764213498851](https://doi.org/10.1177/0002764213498851).
- Song, D. (heinäkuu 2018). ”Learning Analytics as an Educational Research Approach”. *INTERNATIONAL JOURNAL OF MULTIPLE RESEARCH APPROACHES* 10, s. 102–111. DOI: [10.29034/ijmra.v10n1a6](https://doi.org/10.29034/ijmra.v10n1a6).
- Suhonen, S. ja Kinnari-Korpela, H. (2019). ”Using Moodle Data for Early Warning of Dropping Out”. *Varietas delectat... Complexity is the new normality. SEFI 47th Annual Conference Proceedings*, s. 1113–1122. ISSN: 978-2-87352-018-2.
- Toivola, M. (2017). *Flipped learning: käänteinen oppiminen*. 1. painos. Helsinki: Edita. ISBN: 978-951-37-7238-3.
- Wolff, A., Zdrahal, Z., Nikolov, A. ja Pantucek, M. (2013). ”Improving Retention: Predicting at-Risk Students by Analysing Clicking Behaviour in a Virtual Learning Environment”. Teoksessa: *Proceedings of the Third International Conference on Learning Analytics and Knowledge - LAK '13*. Leuven, Belgium: ACM Press, s. 145. ISBN: 978-1-4503-1785-6. DOI: [10.1145/2460296.2460324](https://doi.org/10.1145/2460296.2460324).

Liite A Tekoälyn käyttö tutkielmassa

Tekoälyä ei ole käytetty osana tutkielmaprosessia.