



Kandidatutkielma

Tietojenkäsittelytieteen kandiohjelma

# Oppijan kehittymisen tukeminen oppimisanalytiikalla Moodlessa

Tuomas Alanen

31.5.2022

MATEMAATTIS-LUONNONTIETEELLINEN TIEDEKUNTA  
HELSINGIN YLIOPISTO

## Yhteystiedot

PL 68 (Pietari Kalmin katu 5)  
00014 Helsingin yliopisto

Sähköpostiosoite: [info@cs.helsinki.fi](mailto:info@cs.helsinki.fi)  
URL: <http://www.cs.helsinki.fi/>

HELSINGIN YLIOPISTO – HELSINGFORS UNIVERSITET – UNIVERSITY OF HELSINKI

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			
Prof. D.U. Mind, Dr. O. Why			
Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages	
Kandidutkielma	31.5.2022	19 sivua	
Tiivistelmä — Referat — Abstract			
<p>Kirjoita tiivistelmä tähän.</p> <p>Varmista, että kaikki pakolliset kohdat lomakkeella on täytetty.</p> <p>Listaa ACM CCS 2012 -luokituksesta 1-3 polkua kuvaamaan työtäsi. Kts englanninkielinen ohje.</p>			
<p><b>ACM Computing Classification System (CCS)</b></p> <p>General and reference → Document types → Surveys and overviews</p> <p>Applied computing → Document management and text processing → Document management</p> <p>→ Text editing</p>			
Avainsanat — Nyckelord — Keywords			
learning analytics, Moodle			
Säilytyspaikka — Förvaringsställe — Where deposited			
Helsingin yliopiston kirjasto			
Muita tietoja — övriga uppgifter — Additional information			



# Sisällys

<b>1</b>	<b>Johdanto</b>	<b>1</b>
<b>2</b>	<b>Oppimisen analysoinnin tarpeet</b>	<b>2</b>
<b>3</b>	<b>Datamalli oppijan kehittymisestä</b>	<b>4</b>
3.1	Moodle datalähteenä . . . . .	4
3.2	Yleistetty malli . . . . .	5
3.3	Datamallin rakentamisessa huomioitavaa . . . . .	8
<b>4</b>	<b>Datamallin hyödyntäminen oppimisanalytiikassa</b>	<b>11</b>
4.1	Yksittäiseen oppijaan kohdennetut ehdotukset . . . . .	11
4.2	Opetuksen kehittämiseen kohdennetut ehdotukset . . . . .	12
4.3	Ehdotuksien tulkinnan rajoitteet . . . . .	13
<b>5</b>	<b>Yhteenveto</b>	<b>15</b>
	<b>Lähteet</b>	<b>16</b>



# 1 Johdanto

Tutkimuskysymykset:

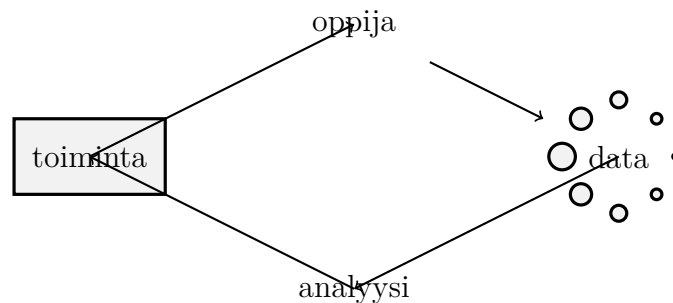
1. Miten oppimisympäriststä saatua dataa voidaan hyödyntää oppijan tukemiseksi oppimisanalytiikan avulla
2. Millaista dataa Moodlesta saadaan oppimisanalytiikan prosessin käyttöön

HOX! Punaisella merkityt tekstiosuudet ovat tutkimuspäiväkirjan sisältöä, johon olen kirjannut ylös erilaisia havaintoja ja hyviä lähdeaineistoja talteen hyödynnettäväksi myöhemmissä vaiheissa. Normaalit tekstiosuudet ovat varsinaista kandidaatin tutkielman sisältöä. Selitettävää: Learning analytics LA  $\leftrightarrow$  Educational Data Mining EMD (Romero ja Ventura, 2010) Virtual Learning Environment VLE  $\leftrightarrow$  Learning Management System LMS LMS on verkkopohjainen järjestelmä, joka mahdollistaa oppimateriaalin, opiskelijoiden toiminnan, tehtävätyökalujen ja oppijan edistymisen seurannan. (Mohd et al., 2016). (Romero et al., 2014)

## 2 Oppimisen analysoinnin tarpeet

Oppimisanalytiikka tutkii oppimista ja opettamista oppimisympäristöissä ja hyödyntää analytiikkaa tunnistessaan poikkeamia tai tehdessään muita oppijaa tukevia havaintoja (Long ja Siemens, 2011). 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 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.

Hallintohenkilöstön toiminnan vaikutukset ovat laajempia ja hitaammin havaittavia heidän yhdistäessä myös opettajalta saatavan palautteen analyysiinsa (Clow, 2012). Hallintohenkilöstö 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.

**Avaa tarpeita vielä lisää.**

# 3 Datamalli oppijan kehittymisestä

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 verkkooppimisympäristö Moodlea.

## 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, ei julkaisupäivää). Moodle on rakennettu käyttäen ohjelmointikielenä PHP:tä 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 aktiviteettiä, 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 aktiveetteja (Moodle.org, 2022). Oppilaan osaamista mittaavia aktiviteetteja 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 taulusta *course\_module\_completion*.

Moodle tallentaa tietokantatauluun *logstore\_standard\_log* 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 saada (Dougiamas, 2022). On teoreettisesti mahdollista, että oppija on avattuaan sivun katsellut sitä minuutin ajan ja tämän jälkeen lähtenyt kahville. Jos seuraava sivulataus on tunnin päästä, niin tästä ei pystytä luotettavasti laskemaan sivulla vietettyä todellista aikaa.

Learning Analytics API tarjoaa Moodlen oman rajapinnan oppimisanalytiikan toteuttamiseen (Olivé et al., 2018). Moodlessa se jakautuu kahteen osaan, Moodlen Analytics API:n ja Machine Learning backendiin. Analytics API tuottaa mallien tarvitsemaa tietoa koneluettavassa CSV-muodossa. Backend puolestaan vastaa itse tiedon käsittelystä ja analysoinnista.

## 3.2 Yleistetty malli

tilastollinen malli kuvaa optimia, ja verrataan kuinka data sopii tähän malliin

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

Oppimisanalytiikassa usein käytetään luokittelua, jota hyödynnetään opetuksessa yleisesti opettajien arvioidessa oppijoiden tietotasoa, motivaatiota ja käytöstä (Hämäläinen ja Vinni, 2010). Oppimisanalytiikassa luokittelua tehdään selitettävän muuttujan arvoa ennustavalla mallilla, jota ennustetaan selittävien muuttujien arvojen avulla. Luokittimia voidaan tehdä joko ammattilaisten käsityönä tai nykyisin yleisemmällä tavalla opettaa luokitin luokittelemaan olemassa olevalla datalla.

Useissa oppimisanalytiikkaa käsittelevissä tutkimuksissa on kokeiltu erilaisia luokittelualgoritmeja parhaiten toimivan mallin löytämiseksi (Akçapınar et al., 2019). Usein käytetyjä algoritmeja ovat naiivi Bayes, Classification Tree, Random Forest, tukivektorikone (SVM), neuroverkko, CN2 rules ja k-lähinaapurimenetelmä. Yksi tapa etsiä parhaiten toimivaa mallia on tehdä suorituskykymittauksia, joissa tarkastellaan tarkkuutta, herkkyyttä, yksityiskohtaisuutta ja F-Measurea.

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 Bayesin 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 oppimisdatalle 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  $\alpha$  ja  $\beta$  ovat tuntemattomia suureita, estimaattoreita, jotka estimoidaan datan perusteella. Muuttuja  $e$  kuvaa satunnaista virhettä, jonka oletetaan noudattavan normaalijakaumaa odotusarvolla 0 ja varianssilla  $\sigma^2$ . Varianssin oletetaan olevan sama riippumatta selittävistä muuttujista  $x$ .

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

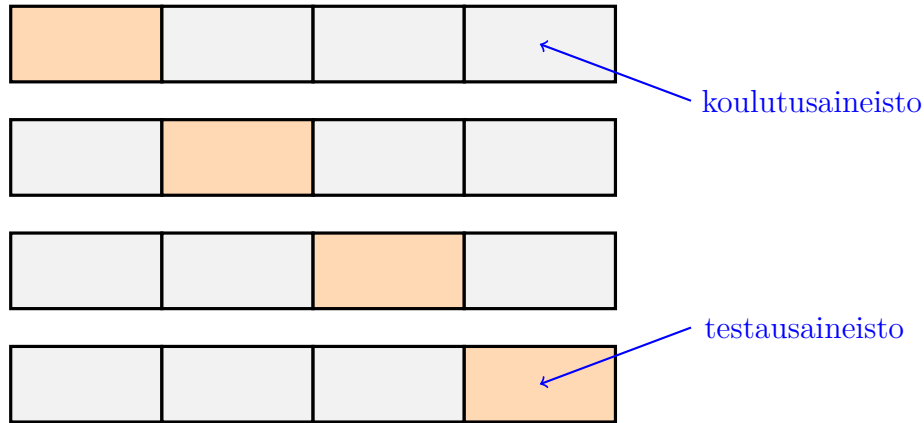
Ennen datan syöttämistä analysointia tai luokittelua tekeväälle mallille, tulee aineistolle suorittaa esikäsittely (Romero et al., 2014). Esikäsittely aloitetaan keräämällä tarvittava data, joka ryhmitellään sopiviin ja järkeviin kokonaisuuksiin. Datan ryhmittelyn jälkeen poistetaan siitä kaikki epäolennainen 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.

Yhdistelläksemme tietyllä välillä liikkuvia muuttujia kategoristen muuttujien kanssa, voidaan käyttää dummy-muuttujia kuvaamaan näitä arvoja (Ross, 2017). Tällöin voidaan hyödyntämään sellaisia selittäviä kategorisia muuttujia, jotka eivät lähtökohtaisesti ole numeerisessa 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  $k$ -suorituskeran muodostamien mallien suorituskyyky 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  $f$  suorituskyykyä 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 ristiinvalidointi 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  $R(f^{(k)}, \mathcal{V}^{(k)})$ .

Useiden eri mallien välisessä vertailussa naiivi Bayes oli ennustustamisen osalta paras algoritmi (Kotsiantis et al., 2004).



## 4 Datamallin hyödyntäminen oppimisanalytiikassa

Oppimisanalytiikan hyödyntäjiä on useissa käyttäjäryhmissä, kuten oppijat, opettajat, kurssikehittäjät (opintoesimies tms), organisaatiot ja hallinto henkilöstö (Romero ja Ventura, 2010). Käyttäjäryhmät hyödyntävät

Alusta kertomalla oppimisanalytiikan useammasta hyödyntämiskerroksesta, oppija, opettaja, koulu, valtakunnallinen. Oppimisanalytiikan avulla voidaan tehdä opetuksen kehtämistä (Romero ja Ventura, 2010). Table 1.

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 ennusteen perusteella katsotaan onko oppija vaarassa olla läpäisemättä kurssia.

### 4.1 Yksitaiseen 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 luokittelulla tai ennustavilla malleilla.

Naiivilla Bayesin luokittimella voidaan esimerkiksi yrittää tunnistaa opiskelijoita, jotka ovat vaarassa saada hylätyn osallistumaltaan kurssilta (Barber ja Sharkey, 2012b). Selittävinä muuttujina oli henkilöön liittyviä taustatietoja, suoritettujen opintopisteiden suhde yritettyihin opintopisteisiin sekä toimintaa verkko-oppimisympäristön keskustelualueella. Selittäville muuttujille oli annettu eri painoarvoja riippuen kurssin viikosta.

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

Yksi oppijan menestymiseen perustuva tarkastelu on onko oppija vaarassa pudota kurssilta (Olivé et al., 2018; Suhonen ja Kinnari-Korpela, 2019). Tarkastellaan oppijan toimintaa verkko-oppimisympäristössä ja yritetään löytää eri merkkejä oppijan putoamisesta kurssilta. Tarkastelua voidaan laajentaa eri kurssien väliseksi (Kinnari-Korpela ja Suhonen, 2020) ja analytiikan löytäessä putoamisvaarassa olevan oppijan esimerkiksi oppijan jättäessä ilmoittautumatta kursseille ja lopulta hiljaisesti jättämällä opinnot kesken, voidaan hänelle tarjota kohdistetusti tukea oppimiseen jo aikaisessa vaiheessa.

Moodleen on sisäänrakennettu Learning Analytics API:n avulla opiskelijoiden tippumisen tunnistamisen tarjoava malli (Olivé et al., 2018; Monllaó et al., 2021). **Avaa Moodle osuutta vielä tarkemmin.**

Yksi taso on hyödyntää oppimisanalytiikkaa 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. **(HOX! Tsiikkaa noi muut kolme muuta Longin nostoa sekä table 1)**

Oppimispolku on ohjeistus, joka 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. Pikapalautetta voidaan hyödyntää itsearvioiden toteuttamiseen ja edelleen analytiikan tukena.

## 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 esimerkiksi uuden kurssin luominen tai nykyisen kehittäminen tukemaan osaamisvajeen paikkaamista. Tämä voi näkyä esimerkiksi oppimateriaalin kehittämisenä, mikäli analytiikka osoittaa tietyn osa-alueen tehtävien menevän muita heikommin, kun saman aikaisesti tiettyä opetusmateriaalin osaa tarkastellaan muita enemmän.

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.

Oppimisanalytiikan avulla voidaan kehittää opetussuunnitelmia sekä tukea hallinnon päätöksiä (Romero ja Ventura, 2010; Long ja Siemens, 2011)

Moodleen on sisäänrakennettu Learning Analytics API:n avulla kurssin opetuksen puutumisen tunnistava malli (Monllaó et al., 2021). [Avaa Moodle osuutta vielä tarkemmin.](#)

Yksi ennustamisen mahdollisuus on tarkastella valmistuuko koulutukseen hakija ennusteen mukaan tavoiteaikataulussa (Barber ja Sharkey, 2012a).

Opetuksen tutkijat (Romero ja Ventura, 2010)

## 4.3 Ehdotuksien tulkinnan rajoitteet

Suunniteltaessa oppimisanalytiikan hyödyntämistä täytyy huomioida eettiset näkökulmat. Eettiseltä näkökulmalta huomioitavaa on esimerkiksi datan omistajuus sekä yksityisyydensuoja (Slade ja Prinsloo, 2013). Lisäksi kontekstin tarkasteleminen on tärkeää, sillä eettiset haasteet ovat erilaiset esimerkiksi resurssien optimoimisen ja oppijan suoriutumisen arvioinnin välillä. Jokaisella

1. etiikka?! (Kaila et al., 2019)
2. laki, henkilötieto? (Hannula, 2017)
3. virhearviot ja model bias
4. mallien ennustuksien paikkaapitävyyden todennäköisyydet - kuinka todennäköisesti ennustuksen tulos pitää paikkansa. voidaanko 72 prosentin todennäköisyyttä pitää

sellaisena, että se toimii luotettavana ohjauksen työkaluna?

## 5 Yhteenveto

Havaintona, että viime vuosina Moodlella tehtyt oppimisanalytiikan tutkimukset ovat vähentyneet ja useat tutkimukset ovat käyttäneet nykymittapuulla vanhentuneita Moodle-versioita.

Oppimisanalytiikalla ei voida korvata oppijoiden ohjausta, vaan oppimisanalytiikka on yksi työkalu kaikkien muiden työkalujen joukossa (Auvinen, 2017). Oppimisanalytiikalla voidaan tehostaa tätä toimintaa.

Useat ennusteet hyödyntävät paljon muutakin dataa kuin oppimisympäristöstä saatavaa dataa

# 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?*
- 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).
- (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).
- (2013). "An Overview of Learning Analytics", s. 23.
- 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*.

- Dougiamas, M. (2021). "The Power of Open Educational Technology". Teoksessa: *Reimagining Digital Learning for Sustainable Development*. Routledge. ISBN: 978-1-00-308969-8.
- (toukokuu 2022). *Moodle*. Moodle HQ.
- 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*.
- 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 »*.
- Green, M. (maaliskuu 2022). *Moodle\_3.11 Database*. <https://www.examulator.com/er/output/index.htm>
- Hannula, H. (maaliskuu 2017). *Oppijan digitaalinen jalanjälki – oppimisen arjessa kertyvät henkilötiedot sekä oikeus ja mahdollisuudet niiden käyttämiseen*.
- 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).
- Kaila, E. T., Kurvinen, E. ja Apiola, M.-V. (2019). "Ethical Considerations in Learning Analytics: Ethics". *CEUR Workshop Proceedings* 2505, s. 61–63. ISSN: 1613-0073.
- Kinnari-Korpela, H. ja Suhonen, S. (heinäkuu 2020). *Oppimisanalytiikalla tehokkaampaan ohjaukseen*. [https://tamkjournal.tamk.fi/oppimisanalytiikalla-tehokkaampaan-ohjaukseen/](https://tamkjournal.tamk.fi/oppimisanalytiikalla-tehokkaampaan-ohjaukseen/publication) publication.
- Kotsiantis, S., Pierrakeas, C. ja Pintelas, P. (toukokuu 2004). "PREDICTING STUDENTS' PERFORMANCE IN DISTANCE LEARNING USING MACHINE LEARNING TECHNIQUES". *Applied Artificial Intelligence* 18.5, s. 411–426. ISSN: 0883-9514, 1087-6545. DOI: [10.1080/08839510490442058](https://doi.org/10.1080/08839510490442058).
- Long, P. ja Siemens, G. (2011). "Penetrating the Fog: Analytics in Learning and Education", s. 6.

Mohd, K., Nurul, N. ja Khalid, F. (2016). *Choosing the Right Learning Management System (LMS) for the Higher Education Institution Context: A Systematic Review - Helka*. <https://helka.helsinki.fi>.

Monllaó, D., Dalton, E., Bösch, L., Aherne, M. ja Mudrak, D. (heinäkuu 2021). *Analytics API - MoodleDocs*. [https://docs.moodle.org/dev/Analytics\\_API](https://docs.moodle.org/dev/Analytics_API).

Moodle.org (marraskuu 2022). *Moodle Plugins Directory - Activity Modules*. <https://moodle.org/plugins/?q>  
– (ei julkaisupäivää). *Moodle Statistics*. <https://stats.moodle.org/>.

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).
- Slade, S. ja Prinsloo, P. (lokakuu 2013). "Learning Analytics: Ethical Issues and Dilemmas". *American Behavioral Scientist* 57.10, s. 1510–1529. ISSN: 0002-7642, 1552-3381. DOI: [10.1177/0002764213479366](https://doi.org/10.1177/0002764213479366).
- 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*. <http://www.theseus.fi/handle/10024/333373>. Publication.
- 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).

