

Arto Pärnänen

**TEKOÄLYN HYÖDYNTÄMINEN AMMATILISESSA
OPETUKSESSA - KOULUTUKSEN JÄRJESTÄJÄN NÄ-
KÖKULMA**



JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2021

TIIVISTELMÄ

Pärnänen, Arto

Tekoälyn hyödyntäminen ammatillisessa opetuksessa – koulutuksen järjestäjän näkökulma

Jyväskylä: Jyväskylän yliopisto, 2021, 72 s.

Tietojärjestelmätiede, pro gradu

Ohjaajat: Abrahamsson, Pekka ja Kultanen, Joni

Tutkielman tavoitteena on selvittää, miten ammatillinen koulutuksen järjestäjä voi hyödyntää tekoälyä omassa opetustoiminnassa. Ammatillisen koulutuksen rahoitusperusteet ovat muuttuneet reformin myötä, sekä myös kokonaisrahoitus on pienentynyt. Tämä on johtanut siihen, että koulutuksen järjestäjät ovat joutuneet tiukentamaan nykyisiä käytössä olevia resurssejaan. Tekoälyn sovellusten katsotaan parantavan organisaatioiden kilpailukykyä, sekä esimerkiksi sen lisäämän automaation tai avustetun päätöksen teon avulla organisaatio pystyy hyödyntämään käytössä olevat resurssit muuhun työhön. Tämän tutkimuksen kirjallisuusosiossa tutustutaan aiempaan tutkimustietoon siitä, mitä data-analytiikan tai tekoälyn sovelluksia koulutuksen järjestäjät voivat hyödyntää ja käydään läpi aiheeseen liittyvää käsitteistöä. Tutkimuksen empiirisessä osiossa perehdytään tapaustutkimuksena Itä-Savon koulutuskuntayhtymän toimintaan, ja selvitetään, kuinka kohdeorganisaatio on pystynyt hyödyntämään tekoälyä ammatillisessa opetuksessa. Tutkimuksessa hyödynnetään Mikalefin, Fjørtoftin ja Torvatnin (2019) viitekehystä tekoälykyvykkyydestä ja kilpailukyvyistä.

Asiasanat: tekoäly, koneoppiminen, oppimisanalytiikka, ammatillinen koulutus

ABSTRACT

Pärnänen, Arto

Utilization of artificial intelligence in vocational education – the perspective of the educational institution

Jyväskylä: University of Jyväskylä, 2021, 72 pp.

Information Systems, Master's Thesis

Supervisors: Abrahamsson, Pekka and Kultanen, Joni

The aim of the dissertation is to find out how a vocational institution can utilize artificial intelligence in their own teaching activities. The funding bases for vocational education have changed with the reform, and total funding has also decreased. This has led to vocational institutions having to tighten their existing available resources. Artificial intelligence applications are considered to improve the competitiveness of organizations, and, for example, through increased automation or assisted decision-making, an organization can utilize the available resources for other work. The literature section of this study examines previous research data on which applications of data analytics or artificial intelligence can be utilized by educational organizations and reviews related concepts. The empirical part of the study examines the activities of the Eastern Savo Municipal Federation of Education as a case-study and examines how the target organization has been able to utilize artificial intelligence in vocational education. The study utilizes Mikalef's, Fjørtoft's and Torvatn's (2019) framework of AI Capabilities and competitive performance.

Keywords: Artificial Intelligence, machine learning, learning analytics, vocational education

KUVIOT

KUVIO 1 Esimerkki päätöspuun loogisista testeistä.....	13
KUVIO 2 Neuroverkko (LeCun ym., 2015).....	14
KUVIO 3 Ihmisen ja tekoälyn yhteistyö "human-in-the-loop" (Grønsund & Aanestad, 2020).....	18
KUVIO 4 Tietovarastoinnin ETL-prosessin kuvaus (Vassiliadis ym., 2002).....	21
KUVIO 5 Tietoaltaan prosessikuvaus (Miloslavskaya & Tolstoy, 2016).....	22
KUVIO 6 Tiedonloughinnan hyödyntäminen koulutuksen järjestäjän systeemeissä (Romero & Ventura, 2007).....	23
KUVIO 7 Yleisimmin käytetyt ominaisuudet opiskelijatietojen tiedonloughinnassa (Alturki ym., 2020).....	25
KUVIO 8 Opiskelijoiden poissaolojen ja kurssimenestyksen vertailu (Lukkarinen ym., 2016).....	26
KUVIO 9 Tekoälykyvykkyyden ja kilpailukyvyn viitekehitys (Mikalef ym., 2019).....	29
KUVIO 10 Sipulimalli (Saunders ym., 2009).....	33
KUVIO 11 Itä-Savon koulutuskuntayhtymän organisaatorakenne (Samiedu, 2021).....	35
KUVIO 12 Samiedun strategiset päätavoitteet ja menestystekijät 2018–2021 (Samiedu, 2021).....	37
KUVIO 13 Samiedun uudistumisen strategian (2018-2021) ja jatkuvan laadunparantamisen prosessit ja työkäytännöt (Samiedu, 2021).....	38
KUVIO 14 Samiedun järjestelmäintegraatiot tietovarastoon.....	43
KUVIO 15 Pelkistetty näkymä tietovaraston rakenteesta.....	45
KUVIO 16 Tekoälymallinnuksen yhteiskunnalliset vaikutukset dokumentoituna ECCOLA-kortilla.....	48
KUVIO 17 Älykäs ohjaus -hankkeessa käytetty tietoarkkitehtuuri.....	50
KUVIO 18 Käyttäjille voidaan esittää heidän omat opiskelijansa käyttäjätunnusten perusteella.....	52
KUVIO 19 Esimerkki vastuuhjaajan työpöydästä, jossa esitetään tekoälyn ennuste vihreänä tai punaisena indikaattorina opiskelijan yksilöivän tiedon alapuolella (sarake 2).....	52
KUVIO 20 Vastuuhjaajan työpöydältä avattavalla opiskelijakohtaisella aikajanakuvaajalla esitetään opiskelijan poissaolo ja suoritekertymiä.....	53
KUVIO 21 Mallinnuksen auditointiin liittyvät vaatimukset ja huomiot.....	54
KUVIO 22 Tekoälyn muodostama ehdotus opiskelijan Wilmassa.....	55

TAULUKOT

TAULUKKO 1 Sekavuusmatriisi (Provost & Kohavi, 1998)	15
TAULUKKO 2 Tietokoneen automaation tasot (Parasurama ym., 2000).....	16
TAULUKKO 3 Oppimisanalytiikan eri tasot (Ferguson 2012).	24
TAULUKKO 4 Määrällisen ja laadullisen tutkimusmenetelmän erot (Hirsjärvi ym., 2015).	33
TAULUKKO 5 Samiedun käytössä olevat tietojärjestelmät ja niiden käyttötarkoitus.	40
TAULUKKO 6 Primus lähdejärjestelmän tuottamia tietoja taulukkomuodossa (CSV).....	44
TAULUKKO 7 Esimerkki, kuinka Primus lähdejärjestelmän tiedot on muokattu tietokantaan ladattavaan muotoon.	44
TAULUKKO 8 Mallinnuksen sekavuusmatriisi.	49
TAULUKKO 9 Mallin tunnusluvut.	49
TAULUKKO 10 Aineiston pohjalta muodostetut empiiriset havainnot.	56
TAULUKKO 11 Empiiristen havaintojen vertaaminen esitettyyn kirjallisuuteen.	59

SISÄLLYS

TIIVISTELMÄ	2
ABSTRACT	3
KUVIOT	4
SISÄLLYS.....	6
1 JOHDANTO.....	8
1.1 Motivaatio.....	8
1.2 Tutkimuskysymys	9
1.3 Työn rakenne.....	10
2 KIRJALLISUUSTARKASTELU	12
2.1 Koneoppiminen.....	12
2.2 Tekoäly	16
2.3 Tiedonlouhinta	19
2.4 Tiedonhallinta	20
2.5 Hyödyntäminen oppilaitoksissa.....	22
3 TUTKIMUKSEN TEOREETTINEN VIITEKEHYS.....	29
3.1 Data ja infrastruktuuri	30
3.2 Tekniset- ja johtamistaidot	30
3.3 Organisaation oppiminen ja tekoälykulttuuri.....	30
3.4 Tekoälyn hyödyntäminen	31
4 TUTKIMUSMENETELMÄT.....	32
4.1 Tutkimusstrategia.....	32
4.2 Case-kuvaus	35
4.3 Aineiston käsittely	39
4.4 Aineiston analysointi.....	39
5 TULOKSET.....	40
5.1 Operatiiviset järjestelmät	40
5.2 Tiedonhallinta	42
5.3 Tekoälymallinnus	47
5.4 Tekoälyn hyödyntäminen	51
5.5 Yhteenveto	56
6 KESKUSTELU.....	57
6.1 Käytännön merkitys	57
6.2 Tutkimuksellinen merkitys	58

7	YHTEENVETO	61
7.1	Vastaukset tutkimuskysymyksiin	63
7.2	Tutkimuksen heikkoudet	64
7.3	Jatkotutkimusaiheet.....	66
8	LÄHTEET	67

1 Johdanto

Koneoppimisen ja tekoälyn sovellusten avulla organisaatiot voivat kehittää omaa toimintaansa resurssitehokkaasti eteenpäin. Sovellukset voivat löytää organisaation tiedoista syvempiä syy-seuraussuhteita tai toistuvuuksia, jotka eivät välttämättä ole niin selkeitä perinteisille tilastollisille menetelmille. Lisäksi nykyaikana tiedon määrä kasvaa jatkuvasti, joten automatisoiduilla menetelmillä tiedon hyödyntämisaste suurenee. Tekoälyn ja koneoppimisen sovelluksilla voidaan toteuttaa automaatiota tai se voi tarjota avustusta päätöksen tekoa varten.

1.1 Motivaatio

Toisen asteen ammatillisessa koulutuksessa toteutetun reformin myötä ammatillisten koulutuksen järjestäjien rahoitusta leikattiin ja painotettiin enemmän suoritusperusteiseksi. Tämän seurauksena koulutuksen järjestäjät ovat joutuneet sopeuttamaan ja tehostamaan omia toimintojaan. Reformin myötä myös opiskelijoiden opintopolut yksilöityivät, eli opinnot voivat edetä enemmän henkilökohtaisen opetussuunnitelman mukaisesti. Tämä asettaa haasteita varsinkin opiskelijan vastuuhjaajalle, sillä yhdellä vastuuhjaajalla voi olla useita kymmeniä opiskelijoita seurattavana, joiden opinnot etenevät yksilölliseen tahtiin.

Koneoppimisen ja tekoälyn sovelluksia voidaan pitää toimintana tehostavina. Ne voivat tuoda informaatiota päätöksen tekoa varten, tai niiden avulla on mahdollista esimerkiksi suorittaa automatisoituja tehtäviä. Tutkimuksen tarkoituksena on selvittää, miten näitä sovelluksia voitaisiin hyödyntää koulutuksen järjestäjän toiminnassa. Koulutuksen järjestäjän operatiiviset järjestelmät tuottavat jatkuvasti uutta aineistoa, jota voidaan hyödyntää erilaisilla analyttikan menetelmillä. Esimerkiksi opetus- ja kulttuuriministeriön oppimisanalyttikan viitekehityksessä todetaan seuraavaa:

Instituutiotasolla tarkastellaan oppijoita, koulutusohjelmia ja opetuksen järjestäjää koskevaa dataa, jota hyödynnetään tietoperustaisen päätöksenteon ja johtamisen tukena. Tietoa kerätään usein oppimisen hallintajärjestelmistä sekä palaute- ja oppilastietojärjestelmistä. Tiedonkeruun tavoitteena on esimerkiksi löytää keinoja koulutuspolkujen sujuvoittamiseen, seurata ja ennakoida tutkintojen valmistumista tai ehkäistä opintojen jäämistä kesken. Instituutiotason analytiikka on työkalu, jonka avulla opetuksen järjestäjät ja niiden ylläpitämät oppilaitokset voivat kehittää esimerkiksi oppijoiden erilaisia tukipalveluita kuten opinto-ohjausta. (Opetus- ja kulttuuriministeriö, 2021)

Tässä kirjallisuuskatsauksessa on tarkoitus perehtyä tarkemmin siihen, miten koneoppimisen ja tekoälyn sovelluksia voidaan hyödyntää käytännössä. Kirjallisuuskatsauksen avulla pyritään saamaan tieteellistä tietoa siitä, kuinka nämä koneoppimisen ja tekoälyn mallit toimivat ja mitä käytännön sovelluksia niiden avulla voidaan tehdä. Tämä kysymys pyritään kohdentamaan opetuksen toimialalle. Kirjallisuusosiossa käydään läpi myös aiheeseen liittyviä käsitteitä, sillä se luo pohjan tutkimukselle. Kirjallisuusosio on osa pro gradu -työtä, jossa tehdään empiirinen tapaustutkimus ammattioppilaitoksen toimintaan, kuinka he hyödyntävät tekoälyä omassa toiminnassaan.

Kirjallisuuskatsauksen sisältö jaetaan kahteen osaan. Ensimmäisessä osassa tutkitaan kirjallisuutta tekoälyyn ja koneoppimiseen liittyen, sekä niihin liitettyihin käsitteisiin. Toisessa osassa keskitytään koneoppimisen ja tekoälyn sovelluksiin käytännössä. Toisen osion sisältö pyritään rakentamaan niin, että se liittyy mahdollisimman vahvasti oppilaitostoimintaan, mutta sovelluksia voidaan tarkastella myös muilta toimialoilta vertailun vuoksi.

1.2 Tutkimuskysymys

Tässä tutkimuksessa selvitetään, miten ammattioppilaitokset pystyvät hyödyntämään tekoälyä omassa toiminnassaan. Tutkimus pyrkii täten vastaamaan seuraavaan tutkimuskysymykseen:

Miten koulutuksen järjestäjä voi hyödyntää tekoälyn ratkaisuja omassa toiminnassaan?

Tutkimuksen päätutkimuskysymys voidaan mieltää kaksiosaiseksi: kuinka koulutuksen järjestäjä saavuttaa tilanteen, että hän voi ottaa käyttöön tekoälynratkaisuja, sekä miten mahdollisia tekoälyn malleja voidaan hyödyntää. Tämän perusteella tutkimukselle voidaan esittää kaksi apututkimuskysymystä:

Kuinka tekoälymallinnus on pystytty rakentamaan?

Kuinka tekoälymallin tuottamaa tietoa hyödynnetään?

Kirjallisuuskatsauksen aineistoksi pyritään valikoimaan mahdollisimman tuoreita aiheeseen liittyviä tutkimuksia. Koneoppimisen ja tekoälyn mallit ovat kuitenkin yleisesti ottaen hyvin tutkittuja, joten näihin liittyvät lähteet saattavat olla hieman vanhempia. Tutkimuksia ei rajata tarkemmin esimerkiksi kielen tai opetusasteen perusteella, sillä oletettavasti oppilaitoksissa hyödynnettäviä tekoälyn tai koneoppimisen sovelluksia on tutkittu Suomessa vähäisesti. Aineistot pyritään valitsemaan niiden sisällön ja julkaisuvuoden perusteella, painottaen uudempia tutkimuksia. Aineistoa etsitään esimerkiksi Google Scholarin ja IT-alan tietokantojen kautta seuraavilla hakusanoilla ja teemoilla:

Machine learning, artificial intelligence, educational data mining, learning analytics, student's performance prediction

Aineistoja pyritään painottamaan koulutussektorilta, mutta yleinen teoria esimerkiksi koneoppimiseen ja tekoälyyn ja niihin liittyviin käsitteisiin voidaan löytää yleisemmillä hakusanoilla. Koulutussektorin aineistot pyritään tarkentamaan liitteellä *education*. Koneoppimiseen ja tekoälyyn liitetään lyhyen tutkielman perusteella termejä *data mining, data analytics, big data* ja *data warehousing*, jotka otetaan myös hakusanoina tarkasteluun.

1.3 Työn rakenne

Tutkimuksen kirjallisuussosiossa pyritään selvittämään, mitä koneoppimisen ja tekoälyn mallinnukset tarkoittavat, ja miten koulutuksenjärjestäjät ovat voineet hyödyntää näitä mallinnuksia käytännössä. Kirjallisuuskatsauksen tavoite on esitellä tutkimukselle pohja, kuinka nämä mallinnukset toimivat käytännössä ja minkälaisia käytännön sovellutuksia koulutuksen järjestäjät pystyvät niiden avulla tekemään ja mitä mahdollisia rajoituksia näiden käyttöönotolla on koulutuksen järjestäjän näkökulmasta.

Tutkimuksen kolmannessa luvussa esitellään Mikalefin, Fjørtoftin ja Torvatnin (2019) teoreettinen viitekehys ”*AI capabilities and competitive performance*”. Viitekehysten tarkoituksena on asettaa lähtökohdat empiirisen tutkimuksen toteuttamiselle, sekä asioille, joihin tutkimuksessa kiinnitetään huomiota. Viitekehys kuvaa ulottuvuudet, joiden avulla organisaatio voi saavuttaa tekoälyvykkyyden sekä hyödyt, joita tekoälyn avulla voidaan saavuttaa liiketoiminnassa.

Tutkimusmenetelmä-kappaleessa esitellään tässä tutkimuksessa käytetty tutkimusmenetelmä. Tutkimuksen tutkimusstrategia on tapaustutkimus, joten kappaleessa esitellään myös kohdeorganisaationa, johon tutkimus on toteutettu. Lisäksi organisaatiosta esitellään heidän organisaatiorakenteensa sekä strategiansa yleisellä tasolla.

Tulokset-luvussa käsitellään toteutetun tutkimuksen aikana syntyneet tulokset. Tulokset esitellään kappaleessa tutkimusasetelmassa luodun kaavan mukaisesti jaoteltuna omiin osioihin. Tutkimuksen tuloksista pyritään esittä-

mään primaareja empiirisiä kontribuutioita (*primary empirical contribution, PEC*), jotka tiivistävät tehdyt empiiriset havainnot. Keskustelu-luvussa pohditaan esille tulleiden tulosten

1. käytännöllistä merkitystä
2. tutkimuksellista merkitystä.

Tulosten tutkimuksellinen merkitys arvioidaan lisäksi kolmesta eri näkökulmasta:

- a) tukee olemassa olevaa tietoa
- b) on uutta tietoa
- c) on ristiriidassa olemassa olevan tiedon kanssa.

Yhteenvedossa tutkimuksen tulokset käydään läpi suhteessa tutkimuksen alussa asetettuun tutkimuskysymykseen. Yhteenvedon tarkoitus on tiivistää tutkimuksen lopputulema, sekä arvioida tutkimuksen tuloksia myös kriittisesti. Yhteenvedossa esitellään myös potentiaalisia jatko- ja lisätutkimusaiheita, joita tämän tutkimuksen pohjalta nousee esille.

2 Kirjallisuustarkastelu

Tässä kappaleessa käydään läpi keskeisimmät käsitteet, jotka liittyvät koneoppimiseen ja tekoälyyn. Kirjallisuustarkasteluun on pyritty ottamaan mukaan käsitteet, jotka ovat olleet oleellisesti esillä tutkimuksissa, jotka ovat liittyneet koneoppimiseen ja tekoälyyn. Lisäksi kirjallisuustarkastelussa esitellään aiempaa tutkimustietoa siitä, miten koulutuksenjärjestäjät ovat pystyneet hyödyntämään data-analytiikan, koneoppimisen ja tekoälyn menetelmiä.

2.1 Koneoppiminen

Koneoppimisella voidaan tarkoittaa esimerkiksi ohjelmistoa, jonka toiminta perustuu aiemmin kerättyyn tietoon ja ohjelmisto kehittyy tämän tiedon pohjalta. Sille ei ole pakko asettaa mitään tiettyä algoritmia, vaan ohjelmisto pystyy toimimaan itsenäisesti. Koneoppimisessa hyödynnetään tilastotiedettä, mutta painottuu enemmän algoritmien kehitykseen, eikä koneoppiminen esitä aineistolle oletuksia sen jakautumisesta. Koneoppimisen eri algoritmeilla voidaan löytää esimerkiksi aineistoista säännönmukaisuuksia, mitä ihminen ei itsenäisesti löytäisi (Shalev-Shwartz & Ben-David, 2014).

Ohjatulla koneoppimisella (*supervised learning*) tarkoitetaan menetelmää, jossa syötetyn datan pohjalta luokitellaan aineistoa. Ohjatussa koneoppimisessa käytetään opetusaineistoa, joka pitää sisällään luokittelutiedon eli ”oikean vastauksen”. Oppimisen jälkeen algoritmi pyrkii luokittelemaan tuntemattoman aineiston (Shalev-Shwartz & Ben-David, 2014).

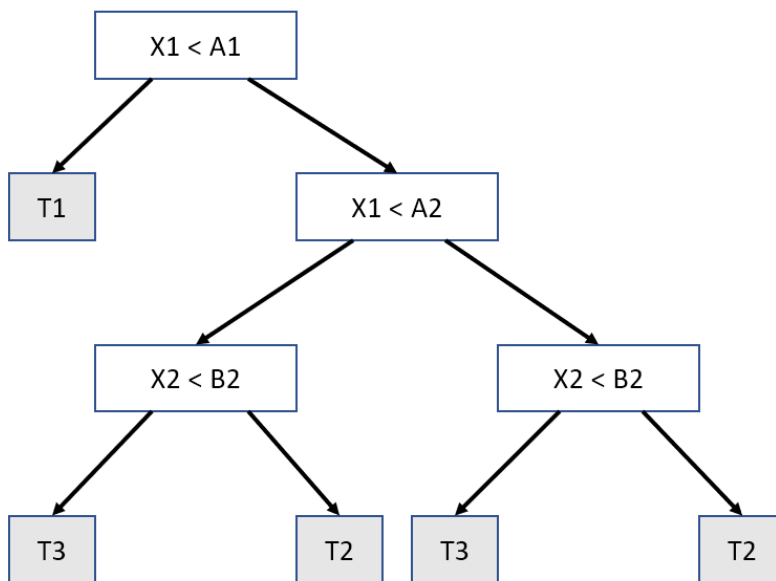
Ohjaamaton oppiminen (*unsupervised learning*) eroaa ohjatusta oppimisesta sillä, ettei aineiston luokkia tunneta ennalta. Aineisto jaetaan luokkiin niiden yhteneväisyyksien perusteella (Shalev-Shwartz & Ben-David, 2014). Klusterointi (*clustering*) on yksi esimerkki ohjaamattomasta oppimisestä. Sen perusteella aineistoista voidaan löytää säännönmukaisuuksia.

Koneoppimisen malleja voidaan rakentaa esimerkiksi päätöspuiden (*decision tree*), tukivektorikoneiden (*support-vector machines*), regressioanalyysin tai

bayesian neuroverkon avulla (Shalev-Shwartz & Ben-David, 2014). Erilaisia koneoppimisen tekniikoita ovat:

- *Päätöspuut* (decision tree), joka on luokittelumenetelmä. Menetelmässä testataan tiedon ominaisuuksia, ja tämän pohjalta tehdään päätös mihin haaraan (branch) edetään. Luokittelun tulosta kutsutaan lehdeksi (leaf), johon saavutaan, kun päätöspuun päätelmät on tehty (Song & Ying, 2015).
- *Tukivektorikone* (support vector machine), joka on luokittelu ja regressionalyysin malli, jossa aineisto analysoidaan tukivektoreiden avulla (Noble, 2006).
- *Naiivi Bayes*, joka on luokittelumenetelmä, mikä olettaa, että havainnot ovat toisistaan ehdollisesti riippumattomia (Yu-Wei, 2015)
- *Neuroverkot*, joka on menetelmä, mikä koostuu useammasta kerroksesta. Edellisen kerroksen tulosta käytetään syötteenä seuraavalle kerrokselle (Li, Jiang, Yang & Wu, 2018)

Päätöspuut koostuvat loogisista testeistä, joiden avulla edetään johtopäätökseen (kuvio 1). Päätöspuiden ongelmaksi voi kuitenkin muodostua niiden koko, jolloin puhutaan ylisovittamisesta. Tällöin malli voi kuvata opetusaineistoa täydellisesti, mutta se ei toimi hyvin aineistolla, jota se ei ole nähnyt. Tätä voidaan estää rajoittamalla päätöspuun kokoa ja käyttämällä useamman päätöspuun kokonaisuutta. Tällöin mallin lopputulos muodostuu siinä olevien päätöspuiden enemmistötuloksen perusteella. (Shalev-Shwartz & Ben-David, 2014).

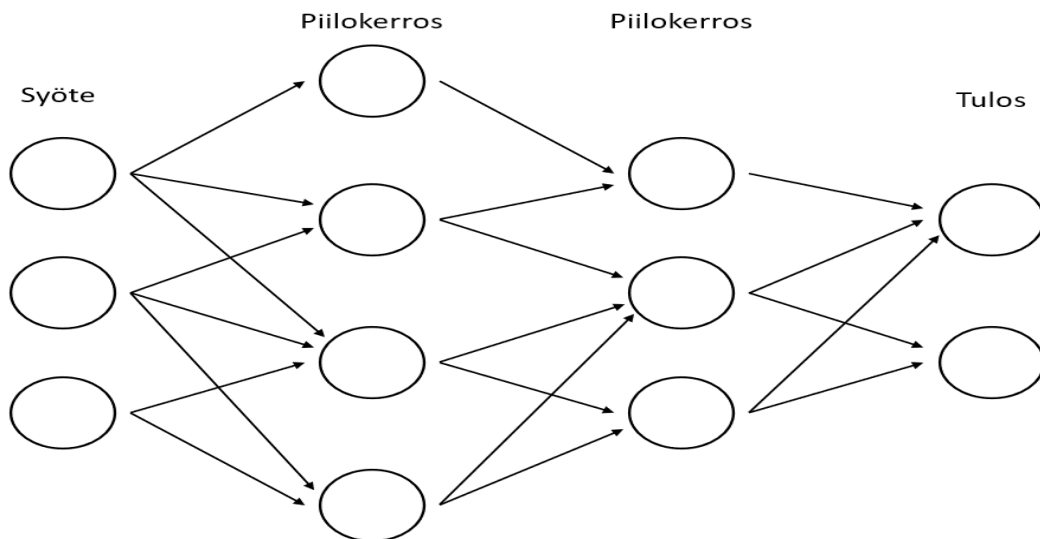


KUVIO 1 Esimerkki päätöspuun loogisista testeistä.

Songin ja Yingin (2015) mukaan päätöspuiden etuja on niiden kyky yksinkertaistaa yhteys syötetietojen ja tulosten välillä jakamalla syöte merkittäviin aliryhmiin. Päätopuut ovat myös ymmärrettäviä ja niitä pystytään tulkitsemaan. Haittapuolena päätöspuilla voi olla niiden ylisovittaminen (*overfitting*) tai alisovittaminen (*underfitting*). Alisovittaminen tarkoittaa liian yksinkertaista päättelyä, ja se voi johtua liian pienestä lähdeaineistosta mallin opettamisessa.

Syväoppiminen on yksi koneoppimisen kategoria (Sengupta, Singh, Leopold, Gulati & Lakshminarayanan, 2020). LeCunin, Bengion ja Hintonin (2015) mukaan syväoppimisen sovelluksilla voidaan löytää monimutkaisia rakenteita suurista aineistoista, sillä ne oppivat tiedosta useilla abstraktiotasoilla ja ne koostuvat useista prosessointikerroksista. Senguptan ja kollegoiden mukaan syväoppiminen käyttää neuroverkkoja oppiakseen uusia asioita.

LeCunin ja kollegoiden (2015) mukaan syväoppiminen perustuu neuroneihin, jotka laskevat syötteen perusteella tuloksen. Syöte voi olla yksittäinen tai koostua useammasta eri lähteestä, jossa arvo voi olla peräisin toiselta neuronilta tai syöttestä. Näistä neuroneista ja niiden kerroksista muodostuu neuroverkko (kuvio 2). Syväoppimisessa tarkoituksena on algoritmien avulla luoda neuroverkko, joka pyrkii ratkaisemaan sille esitetyn ongelman.



KUVIO 2 Neuroverkko (LeCun ym., 2015)

Syväoppiminen on parantanut muun muassa kuvan-, objektien- ja puheentunnistusta. Sen heikkouksiin voidaan kuitenkin lukea mallin vaikea selitettävyyys sekä tarve sen tarve laajalle pohja-aineistolle (LeCun, Bengio, Hinton, 2015).

Provostin ja Kohavin (1998) mukaan koneoppimisen mallien tarkkuutta voidaan mitata aineistolla, jota ei ole käytetty mallin opetusvaiheessa. Garethin (2013) mukaan kerätty havaintoaineisto voidaan jakaa kahteen osaan, opetusaineistoon sekä testausaineistoon. Malli rakennetaan opetusaineistolla, jolloin

testiaineistoon jaetut havainnot eivät ole siinä mukana. Tämän jälkeen mallin tarkkuutta voidaan arvioida sillä, kuinka monta tapausta malli pystyy luokittelemaan oikein testiaineistosta, jonka havaintojen luokittelu on tiedossa. Provostin ja Kohavin mukaan tästä luokittelun testistä voidaan muodostaa sekavuusmatriisi (*confusion matrix*), joka kuvastaa oikein ja väärin luokiteltujen tapausten määrää (taulukko 1). Sekavuusmatriisin koko on $L \times L$, missä L tarkoittaa aineistossa olevien luokkien lukumäärää.

TAULUKKO 1 Sekavuusmatriisi (Provost & Kohavi, 1998)

	Negatiivinen (ennustettu)	Positiivinen (ennustettu)
Negatiivinen (todellinen)	A	B
Positiivinen (todellinen)	C	D

Matriisin avulla pystytään johtamaan tunnuslukuja, joita voidaan käyttää mallin arvioinnissa. Alla on esitetty laskentakaavat mallin tarkkuudesta (*accuracy*), oikeiden positiivisten osuus (*true positive rate*) ja oikeiden negatiivisten osuus (*true negative rate*). Muita hyödynnettyjä tunnuslukuja on myös väärin positiivisten osuus (*false positive rate*) ja väärin negatiivisten osuus (*false negative rate*).

$$accuracy = \frac{A + D}{A + B + C + D}$$

$$true\ positive\ rate = \frac{D}{C + D}$$

$$true\ negative\ rate = \frac{A}{A + B}$$

$$false\ positive\ rate = \frac{B}{A + B}$$

$$false\ negative\ rate = \frac{C}{C + D}$$

Bradleyn (1997) mukaan luokittelun suorituskyvyn mittarina voidaan käyttää ROC-käyrää (*receiver operating characteristics*) ja käyrän alle jäävää pinta-alaa. ROC-käyrä piirretään kuvaajaan, jossa x-akselilla on väärin positiivisten osuus ja y-akselilla oikeiden positiivisten osuus kynnyksarvoittain. AUC (*area under the ROC-curve*) on tämän käyrän alle jäävän osuuden pinta-ala. Bradleyn mukaan AUC-lukuarvo on yksi parhaista yksittäisistä mittareista kuvaamaan luokittelijan suorituskykyä.

2.2 Tekoäly

Tekoälyksi mielletään kone tai ohjelmisto, joka tekee toimintoja, joita voidaan pitää ihmisen kaltaisena älyllisenä toimintana. Tämmöisiä toiminnallisuuksia on esimerkiksi hahmojen tunnistaminen kuvista. Kaplanin ja Haenleinin (2019) mukaan tekoäly määritellään kyvyksi tulkita ulkoisia tietoja oikein ja oppia tällaisista tiedoista. Zhengin ja kollegoiden (2017) mukaan tekoälyn menetelmät ovat kehittyneet viime vuosina voimakkaasti johtuen laskentatehon, tiedon hankinnan ja sen varastoinnin kehittymisen takia. Shalev-Shwartzin ja Ben-Davidin (2014) mukaan koneoppiminen voidaan nähdä yhtenä tekoälyn osa-alueena. Suurin osa tekoälyn sovelluksista on niin sanottuja heikon tekoälyn sovelluksia, jotka pystyvät ratkaisemaan yhden ongelman. Vahva tekoäly puolestaan kykenisi ratkaisemaan useampia ongelmia (Siau & Yang, 2017)

Tutkimusten mukaan tekoälyn käyttöönotto saattaa olla merkittävä yritysten kilpailutekijä (Davenport & Ronanki, 2018). Tekoälyn ajatellaan tuovan lisäarvoa suorittamalla manuaalisia tehtäviä nopeammin, tarkemmin ja paremmin verrattuna ihmisiin. Tekoälyn suorittaessa näitä tehtäviä, ihmisen työaika jää muille tehtäville, joissa ihminen on edelleen parempi (Brynjolfsson, Rock & Syverson, 2018). Vaikka tekoälyn ajatellaan yleisesti korvaavan ihmisen joissain tehtävissä, voi ihmisen ja tekoälyn rooli olla myös yhteinen tehtäviä ratkaistaessa. Tällöin voidaan hyödyntää ihmisten sekä koneiden parhaita puolia (Jarrahi 2018). Mikalefin ja kollegoiden (2019) mukaan tekoälyn hyödyntäminen liiketoiminnassa voidaan jakaa neljään eri pääkategoriaan, joita ovat automaatio, päätöksen teon tukeminen, markkinointi, innovointi.

Jarrahin (2018) mukaan tekoäly voi suuremmalla laskennallisella tiedonkäsittelykapasiteetilla ja analyyttisellä lähestymistavalla laajentaa ihmisten kognitiota käsitellessään monimutkaisuutta. Ihmiset taas pystyvät tarjoamaan kokonaisvaltaisemman ja intuitiivisemmän lähestymistavan epävarmuuden ja epäselvyyden käsittelemiseen.

Araujon, Helbergerin, Kruikemeierin ja De Vreesen (2020) tutkimuksen mukaan joissain tilanteissa tekoälyn tekemät automaattiset päätökset saattoivat olla yhtä hyviä tai parempia kuin ihmisen tekemät päätökset. Automaatiolla tarkoitetaan ihmisen työn tai toiminnan korvaamista. Parasuraman, Sheridanin ja Wickensenin (2000) mukaan tietokoneen automaation tasot jakautuvat kymmeneen tasoon (taulukko 2), joista taso 10 on korkein automaation taso. Heikoimmalla automaation tasolla kone ei avusta ihmistä millään tavalla, vaan ihminen tekee kaikki päätökset itsenäisesti.

TAULUKKO 2 Tietokoneen automaation tasot (Parasurama ym., 2000).

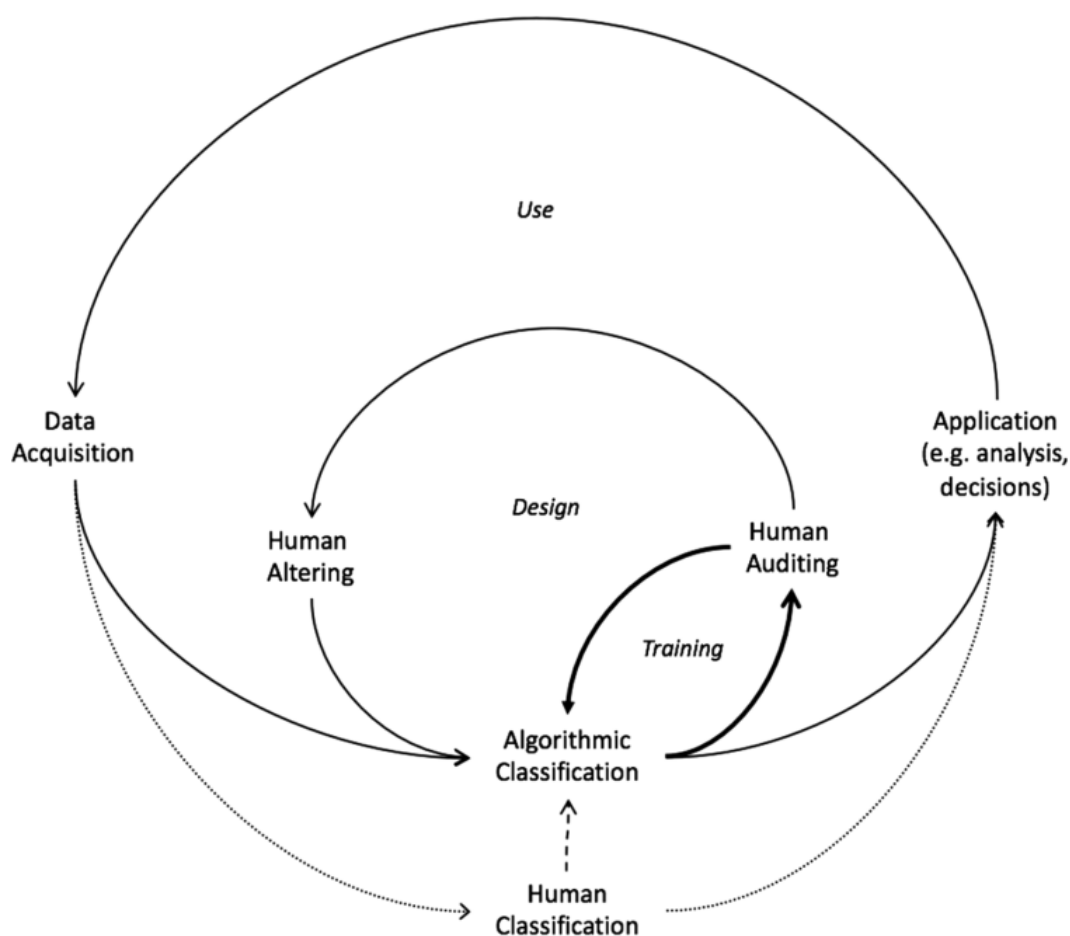
Taso	Tietokone
10	.. päättää kaiken ja käyttäytyy autonomisesti (ihmistä ei huomioida)
9	.. informoi ihmistä, jos päättää niin
8	.. informoi ihmistä, jos siltä kysytään

7	.. toimii automaattisesti, jonka jälkeen informoi ihmistä
6	.. antaa ihmiselle veto-oikeuden rajoitetulle ajalle
5	.. odottaa ihmisen hyväksyntää
4	.. ehdottaa vaihtoehtoa
3	.. rajaa vaihtoehdot muutamaaan
2	.. tarjoaa kaikki vaihtoehdot päätöksentekoon tai toimenpiteeseen
1	.. ei avusta ihmistä, ihminen tekee kaikki päätökset

Mikalefin ja kollegoiden (2019) mukaan tekoälyn automatisoimia tehtäviä voi olla esimerkiksi keskustelubotit (*chatbotit*), jotka kommunikoivat asiakkaiden kanssa. Automatisoitu tekoäly voi myös tarkastella raportteja, dokumentteja ja taloustietoja, minkä ansiosta yrityksen tehokkuus voi parantua.

Tekoäly voi toimia myös ihmisen päätöksen teon tukena. Tätä voidaan kutsua lisätyksi älykkyydeksi (*augmented intelligence*) (Zheng ym., 2017). Zhengin ja kollegoiden (2017) mukaan epävarmuus ja haavoittuvuudet ihmisten elämässä sekä ongelmien monimutkaisuuden takia tekoäly ei pysty kokonaisuudessaan korvaamaan ihmistä. Tällöin tekoälyä voidaan hyödyntää ongelmien ratkaisussa ihmisten tukena. Tekoäly voi tarjota ihmiselle analyyseja, neuvoja sekä täytäntöönpanotukea (Shortliffe & Sepúlveda, 2018). Tällöin ihminen tekee viime kädessä päätökset ja toimenpiteet, mutta voi hyödyntää tekoälyn tuomaa lisäarvoa.

Zhengin ja kollegoiden (2017) mukaan ihmisen ja tekoälyn yhteistyö voidaan jakaa kahteen eri perusmalliin, joista ensimmäisessä ihminen voi olla osana tekoälyn toimintaa (*human-in-the-loop*). Termillä tarkoitetaan ihmisen ja tekoälyn tekemää yhteistyötä, kuten ihmisen täydentämiä puutteita opetusaineistosta tai automatisoitujen järjestelmien ylläpitoa ja kontrollointia esimerkiksi eettisestä näkökulmasta (Grønsund & Aanestad, 2020). Grønsundin ja Aanestadin mukaan ihmisen rooli on esimerkiksi luokittelualgoritmien opettaminen, auditointi ja muokkaaminen (kuvio 3).



KUVIO 3 Ihmisen ja tekoälyn yhteistyö "human-in-the-loop" (Grønsund & Aanestad, 2020)

Zhengin ja kollegoiden (2017) mukaan toinen tekoälyn ja ihmisen yhteistyömalli on kognitiiviseen laskentaan perustuva avustettu älykkyys, jossa kognitiivinen malli on upotettu osaksi koneoppimisen järjestelmää.

Markkinoinnissa tekoälyn sovelluksilla voidaan toteuttaa räätälöityjä ja yksilökohtaisia markkinointikampanjoita, kuten tarjoamalla henkilökohtaisia alennuksia, etuuksia tai palveluita. Tekoälyn käytettävissä voi olla tarkempaa ja hienojakoisempaa tietoa käyttäjästä, joten se voi lähestyä heitä eri tavoin ja yksilöllisemmin. Tekoälyn ja koneoppimisen mallien avulla mainokset voidaan kohdentaa yrityksen tavoittelemalle kohderyhmällä tarkemmin. Tällöin tekoäly korvaa ihmisen markkinoinnissa kehittämällä sopivimman lähestymistavan (Sterne, 2017).

Heerin (2019) mukaan tekoälyä voidaan hyödyntää myös uusien innovaatioiden kehityksessä. Tekoäly mielletään yleisesti pelkästään automaation välineenä, mutta se voitaisiin nähdä työkaluna, joka auttaa ja rikastaa ihmisten älyllistä työtä, eikä niinkään korvaa sitä. Esimerkiksi ammattilaiset voivat hyödyntää tekoälyn tuottamaa syötettä kehittäessään uusia tuotteita tai palveluita.

Vakkurin, Kemellin, Jantusen, Halmen ja Abrahamssonin (2021) mukaan tekoälyn sovellukset yleistyvät ja niillä on kasvava vaikutus yhteiskunnallisesti. Järjestelmien käyttö on tuonut esiin esimerkiksi tietosuojaan liittyviä huolia,

jonka takia tekoälyjärjestelmien eettisyyteen on kiinnitetty enemmän huomiota. ECCOLA-menetelmä esittää toimintatavan, jonka avulla tekoälyjärjestelmien kehittäjät ja niitä kehittävät organisaatiot pystyvät muuttamaan periaatteet ja arvot järjestelmän vaatimuksiksi.

Myös Bostrom ja Yudkowsky (2014) nostavat esille tekoälysovellusten eettisen näkökulman. Heidän mukaansa tekoälyn tekemien päätösten tai toimenpiteiden läpinäkyvyyttä tulisi korostaa. Esimerkiksi tilanne, jossa tekoälyn tekemä päätös vaikuttaa ihmiseen, tulisi olla todennettavissa, miten kone on päätenyt kyseiseen päätökseen. Läpinäkyvyyden lisäksi tekoälyn sovellusten pitäisi olla vaikeasti manipuloitavia. Myös sovelluksen tekemien päätösten tulisi olla ennustettavissa, eli päätökset ovat samansuuntaisia ja linjassa aiempien vastaavien tapausten kanssa.

2.3 Tiedonloughinta

Tiedonloughinnalla pyritään löytämään suurista ja monimutkaisista aineistoista rakenteita ja säännönmukaisuuksia. Tiedonloughinta jaetaan mallin rakentamiseen sekä kaavojen tunnistamiseen. Mallin rakentamisessa on samankaltaisuuksia tilastolliseen mallintamiseen. Kaavojen tunnistaminen pyrkii löytämään aineistoista poikkeamia tai pienempiä rakenteita suuren osan aineistosta ollessa epärelevanttia. (Hand, 2007).

Larosen ja Larosen (2014) mukaan tiedon määrä kasvaa yrityksissä 40 % vuosittain. Tiedon määrän kasvaessa yritykset voivat parantaa tulostaan tai saavuttaa kustannussäästöjä hyödyntämällä tätä tietoa. Tiedonloughintaa kutsutaan prosessiksi, jossa etsitään hyödyllisiä kaavoja ja trendejä suurista tietomassoista. Sen avulla voidaan pyrkiä ymmärtämään palvelun käyttäjän käyttäytymistä, parantaa palvelun tasoa sekä lisätä liiketoimintamahdollisuuksia (Chen, Han & Yu, 1996). Tiedonloughinnan avulla voidaan toteuttaa erilaisia ennustemalleja, esimerkiksi petoksiin, markkinointiin, asiakashankintaan, kysyntään ja talouteen liittyen (Pyle, 1999).

Pyle (1999) kuvaa tiedon muodostumista reaali maailmasta tapahtumina. Tietoa voidaan kuvata koonniksi havainnoista, mitkä koostuvat asioista ja asioista mitkä tapahtuvat ensimmäisen asian jälkeen. Tätä kutsutaan kausaliteetiksi, eli syy-seuraussuhteeksi, jossa yksi asia aiheuttaa toisen (*syy* ja *seuraus*). Tiedonloughinnan menetelmiä voidaan löytää näitä suhteita tietomassasta.

Larosen ja Larosen (2014) mukaan tiedonloughinta voi johtaa myös virheellisiin johtopäätöksiin väärinkäytettyjen menetelmien takia. Tämä voi johtua esimerkiksi analyysistä, joka tehdään väärän lähtötiedon perusteella. Myös tiedonloughinnan mallit voidaan rakentaa täysin väärillä olettamuksilla. Pyle (1999) korostaa myös, että tiedonloughinta vaatii mallissa käytettävän tiedon esivalmistelun (*preparation*). Tiedonloughinnassa tarvitaan ihmisten ohjausta, eikä tiedonloughinta ole välttämättä täysin automaattinen prosessi tai tuote, minkä voi ostaa ja ottaa suoraan käyttöön. Ihmisen rooli tiedonloughinnassa on tutkia ja analysoida aineistoa (Larose & Larose, 2014).

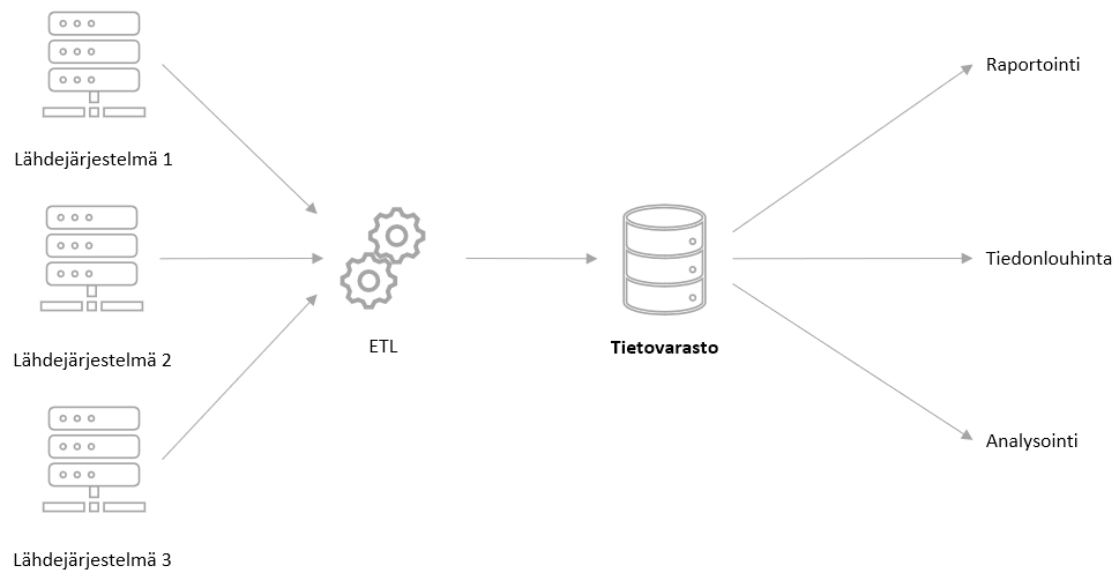
Provostin ja Kohavin (1998) mukaan tiedonlouhinta käsitteenä on ylikuormitettu, sillä joskus tiedonlouhinnalla viitataan tiedon löytämisen prosessiin kokonaisuudessaan. Joskus käsitettä taas saatetaan käyttää viittaamaan tiettyyn koneoppimisen vaiheeseen.

2.4 Tiedonhallinta

Koneoppimisen, tiedonlouhinnan ja erilaisten tekoälysovellusten toteuttamiseksi tarvitaan yleisesti ottaen tietoa (*dataa*). Organisaation tietoja voidaan hallita useammalla eri tekniikoilla, joita voi olla esimerkiksi tietovarastointi (*data warehousing*). Tietovarastoinnin avulla tiedot ovat helpommin ja kattavammin käsiteltävissä, ja sen avulla tiedonlouhinta voi onnistua paremmin. Tämä johtuu tietovaraston rakenteesta, jossa voidaan säilyttää historia-, meta-, yksityiskohtaista- ja summatietoa (Inmon, 1996). Tietovarastossa aineisto on rakenteellisessa muodossa tallennettu.

Kimballin ja Rossin (2011) mukaan tieto on yksi organisaation tärkeimmistä omaisuuksista. Tietoa voidaan säilöä organisaatioiden operatiivisissa järjestelmissä sekä tietovarastosta. Operatiiviset järjestelmät ovat niitä, joihin tietoa syötetään, esimerkiksi uusien tilausten tekemistä, uusien asiakkuuksien perustamista tai palautteen keräämiseen. Niiden tarkoitus on toistaa näitä toimenpiteitä uudelleen ja uudelleen. Tietovaraston tarkoituksena on taas toisaalta luoda näkymä tähän syötettyyn informaatioon, kuten laskemalla uusien tilausten lukumääriä sekä vertailemalla niitä aiemman viikon toteumaan tai tarkastelemalla asiakkaiden tekemiä valituksia. Tietovaraston käyttäjät harvoin tutkivat yksittäisiä tietorivejä, vaan niiden muodostamia kokonaisuuksia. Tietovaraston käyttäjien tarpeet myös muuttuvat useasti.

Vaismanin ja Zimányin (2014) mukaan tietovarastoja voidaan käyttää päätöksentekoprosessin tukena, esimerkiksi strategista tietoa analysoidessa. Tietovarastot keräävät suuria määriä tietoa, ja kokoavat ne sellaiseen muotoon, että sitä voidaan käyttää kuvaamaan organisaation toimintaa. Tietovarastoon tiedot päätyvät erilaisten välivaiheiden kautta: tiedon hakeminen lähteiltä (*extraction*), muokkaaminen (*transformation*), yhdistäminen (*integration*), puhdistaminen (*cleansing*) ja tallentaminen (*storing*). Tietojen siirtämistä lähdejärjestelmistä tietovarastoon kutsutaan myös ETL-operaatioksi (*extraction-transformation-loading*) (Vassiliadis, Simitsis & Skiadopoulos, 2002). ETL-operaatio on havainnollistettu alla olevassa kuviossa (kuvio 4).

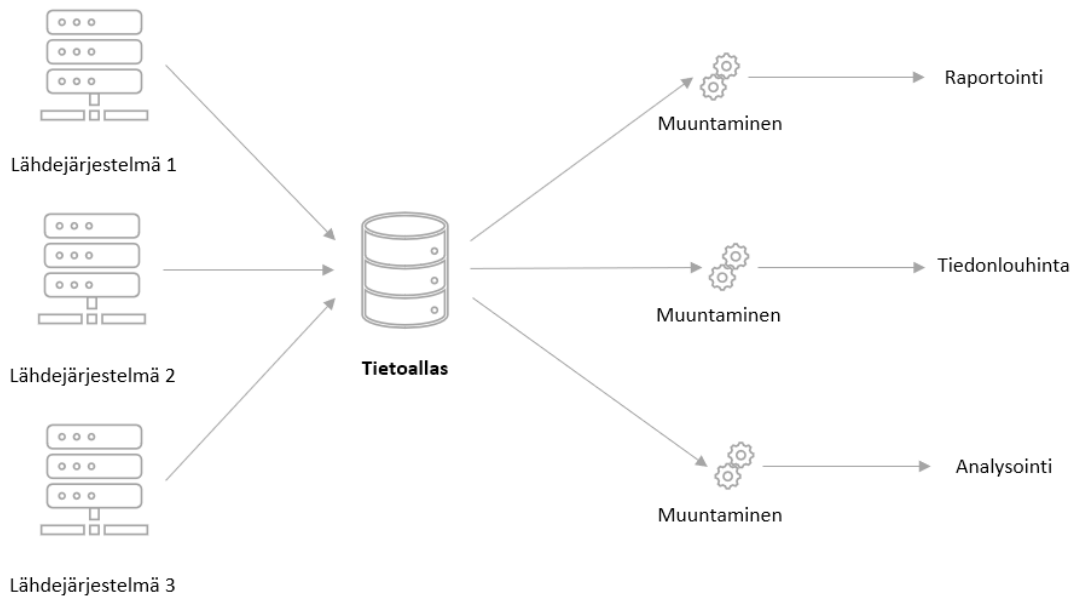


KUVIO 4 Tietovarastoinnin ETL-prosessin kuvaus (Vassiliadis ym., 2002).

Watson, Goodhue ja Wixom (2002) kuvaa tietovarastoprojekteja riskialttiiksi mutta myös erittäin hyödyllisiksi projektin onnistuttua. Tietovarastoinnin tarkoituksena on liiketoiminnan näkökulmasta parantaa päätöksen tekoa sekä organisaation suorituskykyä. Tietovarastoinnin hyödyiksi mainitaan esimerkiksi tiedon tarkkuus, hyödynnettävyys ja luotettavuus.

Miloslavskayan ja Tolstoy (2016) mukaan viime vuosikymmenenä yritysten päätöstentekoon käytettävä tietomäärä on kasvanut valtavasti. Yrityksillä on tarve käsitellä tätä tietoa tunnistukseen oleellinen liiketoimintaan liittyvä tieto. Suurin osa tästä tiedosta on kuitenkin rakenteetonta. Tätä tietoa voidaan käsitellä "big data" konseptilla, jonka periaatteiden mukaan tieto voi olla rakenteellista, puolirakenteellista tai rakenteetonta. Tällaista tietoa ei voida käsitellä perinteisillä tietokannoilla ja sovelluksilla, mutta sitä voi käsitellä tiedonlouhinnan työkalujen avulla. Big datan ja perinteisen tiedon ero voidaan jaotella kolmen V:n mallilla: *volyyymi* (määrä), *velocity* (vauhti) ja *variety* (vaihtelu). Big datassa tiedon määrä ja vauhti on suurta, sekä sen laatu ja rakenne vaihtelee paljon. Myöhemmin on otettu käyttöön myös neljän V:n malli: *veracity* (totuudenmukaisuus), *variability* (vaihtelevuus), *value* (arvo) ja *visibility* (näkyvyys) (Miloslavskaya, Senatorov, Tolstoy & Zapechnikov, 2014).

Tietoaltaan (*data lake*) avulla voidaan varastoida suuria määriä raakatietoa, siinä muodossa kuin se on muodostunut. Tietoaltaaseen tallennettua tietoa voidaan hyödyntää myöhemmin käytön yhteydessä, jolloin raaka-aineistosta voidaan poimia hyödynnettävä tieto ilman tietoaltaan rakenteellista muokkaamista (kuvio 5). Tietoaltaasta haettavat tiedot ja analyysit ovat siis dynaamisia. Tämä poikkeaa tietovarastoon tehtävästä analytiikasta, sillä sen rakenne ja skeema on tiedossa. Tietoaltaassa tallennetut tiedot eivät ole rakenteellista, ennen sen kuin sitä aletaan hakea (Miloslavskaya & Tolstoy, 2016).



KUVIO 5 Tietoaltan prosessikuvaus (Miloslavskaya & Tolstoy, 2016).

Inmonin (2016) mukaan tietoaltaiden ongelmaksi voi muodostua se, että aineistoa kerrytetään, mutta sitä ei pystytä hyödyntämään. Tietoaltaista voi muodostua yhdensuuntaisia, jolloin pelkästään altaan tietomäärä kasvaa jatkuvasti. Syy tähän voi olla esimerkiksi liian epämääräisen tiedon tallentaminen, jolla ei ole yritykselle mitään käyttötarkoitusta. Myös tietojen väliset suhteet voivat hävitä tiedon tallennuksessa. Tällöin oleellisen tiedon löytäminen tietoaltaasta on hankalampaa eikä sitä voida hyödyntää.

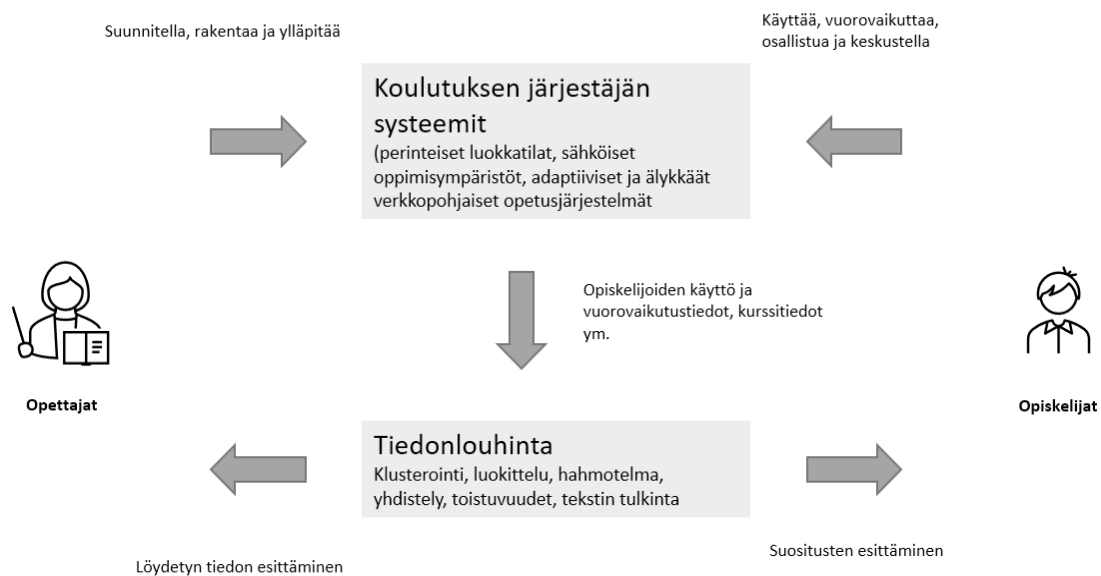
Tietoaltaiden ja "big datan" lisäksi on muodostunut myös "fast data" malli, jota voi pitää näiden kahden menetelmän yhdistelmänä. Fast data pyrkii hyödyntämään tietoa nopeasti ja ratkaisemaan tietyn ongelman (Miloslavskaya & Tolstoy, 2016). Bailisin ja kollegoiden (2017) mukaan fast data -sovelluksia voidaan hyödyntää esimerkiksi ympäristöissä, jotka vaativat suurta suorituskykyä. Tällaisia ovat esimerkiksi telemetriikkayhtiöiden sovellukset, jotka tarkkailevat satoja tuhansia kulkuvälineitä.

2.5 Hyödyntäminen oppilaitoksissa

Tiedonlouhinnan tekniikoiden hyödyntämistä opiskelijatietoihin kutsutaan termillä EDM (*educational data mining*) (Baker & Yacef, 2009). Tällä voidaan tarkoittaa tekniikoiden ja työkalujen käyttämistä tietojen automaattiseen keräämiseen tietovarannoista, joihin kertyy aineistoa opiskelijoiden toimintaympäristöstä (Nithya, Umamaheswari & Umadevi, 2016).

Tiedonlouhinnan avulla voidaan löytää esimerkiksi informaatiota ja tekijöitä, jotka ovat yhteydessä opiskelijasuorituksiin (Yassein, Gaffer, Helali & Mohomad, 2017).

EDM pitää sisällään useita erilaisia tekniikoita, kuten koneoppimista, tilastointia, informaation visualisointia ja tietokone-mallinnusta (Romero & Ventura, 2007). Alla olevassa kuvassa on esitetty esimerkkikaavio tiedonlouhinnan hyödyntämisestä koulutuksen järjestäjän toiminnassa (kuvio 6). Tiedonlouhinta kuvataan iteratiivisena syklinä, jossa systeemiä voidaan myös parantaa aiempien tiedonlouhinnan tulosten perusteella. Systeemiin kerääntyy tietoa jatkuvasti sekä opettajan että opiskelijan näkökulmasta, jota hyödynnetään tiedonlouhinnan menetelmillä.



KUVIO 6 Tiedonlouhinnan hyödyntäminen koulutuksen järjestäjän systeemeissä (Romero & Ventura, 2007)

Oppimisanalytiikalla tarkoitetaan koulutukseen ja oppimiseen liittyvän tiedon keräämistä, analysointia ja raportointia (Larusson & White, 2014). Analytiikassa hyödynnetään digitaalisiin sovelluksiin ja ympäristöihin jääviä digitaalisia jälkiä. Sen tavoitteena on kehittää ja tukea oppimisprosesseja tarjoamalla tietoa päätöksen teon tueksi. Oppimisanalytiikkaa voidaan hyödyntää esimerkiksi antamalla opettajalle tilastotietoa opiskelijoiden toiminnasta visuaalisesti tai informoimalla opettajaa opiskelijoiden kohtaamista haasteista (Baker & Inventado, 2014).

Ferguson (2012) jakaa oppimisanalytiikan neljään eri tasoon: opiskelija, opettaja, instituutio sekä kansallinen. Jokaisella tasolla voidaan määrittää erilaiset tarpeet kerättävän datan laatuun, määrään ja tarkkuuteen. Alla olevassa taulukossa on kuvattu eri tasojen avulla oppimisanalytiikan roolia ja merkitystä (taulukko 3).

TAULUKKO 3 Oppimisanalytiikan eri tasot (Ferguson 2012).

Oppimisanalytiikan taso	Merkitys tai käyttötarkoitus
Opiskelija	Opiskelijan tukeminen opinnoissa
Opettaja	Opetuksen kehittäminen Opiskelijoiden opintojen etenemisen seuranta
Instituutio	Päätöksenteko Johtaminen Ennakointi
Kansallinen	Oman toiminnan kehittäminen Koulutuspoliittinen päätöksenteko Vertailutiedot opetuksen järjestäjien välillä

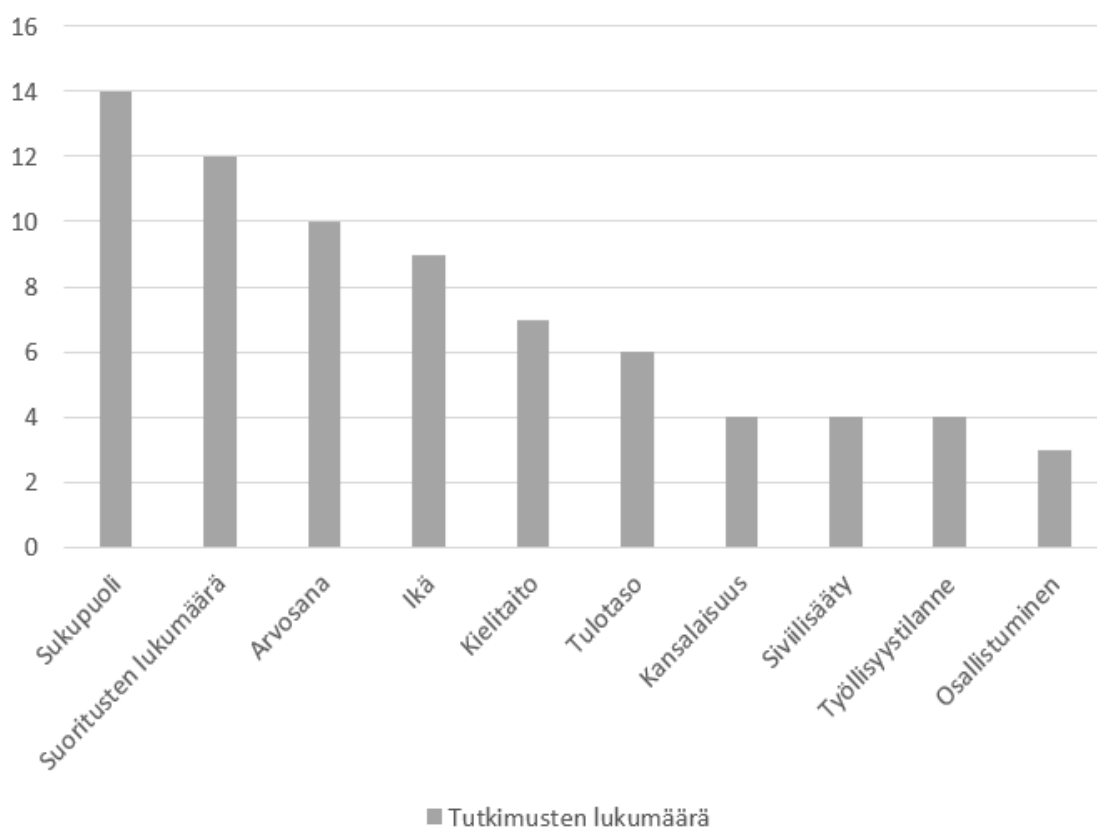
Romeron ja Venturan (2013) mukaan koulutuksen järjestäjän järjestelmät pitävät sisällään suuria tietomääriä useista eri tietolähteistä. Opiskelijatiedonlouhinnalla voidaan ymmärtää paremmin, miten opiskelijat oppivat ja voidaan tunnistaa tekijöitä, mitkä vaikuttavat parempiin opintosuorituksiin.

Koedingerin (2015) mukaan opiskelijatiedonlouhinnan käyttämät tiedot voivat tulla opiskelijoiden vuorovaikutuksesta järjestelmien kanssa. Tiedot voivat olla tarkkaa aikaleimattua tapahtumatietoa oppilaitoksen järjestelmistä tai erilaisilta sensoreilta, jotka voivat havaita opiskelijan silmien ja kehon liikkeitä sekä kasvojen ilmeet. Tietojen perusteella on voitu tehdä erilaisia ratkaisuja esimerkiksi parantamaan opiskelijoiden ohjeistuksia, tarkentamaan tutorointiin liittyvää keskustelua sekä automatisoida työkaluja, jotka avustavat opiskelijan työskentelyä verkkoalustoilla.

Alturki, Hulpus ja Stuckenschmidt (2020) tutkivat opintomenestykseen liittyviä tekijöitä akateemisessa koulutuksessa tiedonlouhinnan avulla. Tutkimus nostaa esille kolme vaihetta, joiden avulla saadaan rakennettua luotettava malli, jossa virheiden todennäköisyys saadaan pienennettyä:

1. oikeiden ominaisuuksien valinta
2. oikean metodin tai tekniikan valinta
3. sopivan tiedonlouhinnan välineen valitseminen

Vaihe 1. jakaantuu demograafisiin tietoihin, opintoja edeltäviin tietoihin sekä opintojen aloitusvaiheen jälkeisiin tietoihin. Demograafiset tiedot koostuvat esimerkiksi sukupuolesta, iästä, perhestatuksesta, tuloista ja vanhempien koulutustasosta. Opintoja edeltävät tiedot koostuvat esimerkiksi aiempien suoritusten keskiarvoista, aiemmasta koulutuksesta tai muista esivalintakokeista. Opintojen aloitusvaiheen jälkeiset tiedot kuvaavat opiskelijan asennetta aloittamisen jälkeen. Tätä voidaan mitata esimerkiksi osallistumisella, suorituksilla, arvosanoilla ja koetuloksilla. Kaaviossa on esitetty tutkimuksessa yleisimmiksi havaitut ominaisuudet, joita on ennusteiden tekemiseen (kuvio 7).



KUVIO 7 Yleisimmin käytetyt ominaisuudet opiskelijatietojen tiedonlouhinnassa (Alturki ym., 2020).

Opiskelijan ikä ja sukupuoli ovat tutkimuksen yksiä eniten käytetyistä ominaisuuksista ennusteita tehdessä. Aineiston perusteella ei kuitenkaan voida sanoa, että nämä olisivat merkittäviä ominaisuuksia ennusteen tekemiselle (Alturki ym., 2020) ja esimerkiksi Kovačićin (2010) tutkimuksen mukaan iällä ei ollut merkitystä opiskelumenestyksen ennustamisen kannalta. Yessin (2009) mukaan opiskelijan siviilisäädyltä on merkitystä opiskelumenestykseen. Man ja Woosterin (2009) mukaan myös avioliitossa olevat opiskelijat suoriutuivat paremmilla arvosanoilla aviottomiin verrattuna.

Opiskelijan työelämän tiedoista (työllisyys sekä tulotaso) on löydetty yhteyksiä opintomenestyksen ennusteeseen. Alin, Haiderin, Munirin, Khanin ja Ahmedin (2013) mukaan opiskelijan tulotaso vaikuttaa merkittävästi opintomenestykseen. Mohamadian, Fallahin, Safdarianin ja Jalalin (2015) tutkimuksen mukaan opiskelijan työllisyydellä on negatiivinen vaikutus opintomenestykseen. Tutkimuksessa arvioidaan, että työssäkäyvät opiskelijat käyttävät vähemmän aikaa opintoihinsa.

Oikeiden metodien ja tekniikoiden valinta on aina tapauskohtaista. Tekniikat voidaan valita testaamalla useampaa eri menetelmää, ja valitsemalla näistä parhaimman lopputuloksen tuottava menetelmä. Päätöspuut ovat yleisin koneoppimisen ja tiedonlouhinnan menetelmä, sillä niiden käyttö tukee käytettävyyttä sekä tehokkuutta. (Alturki ym., 2020).

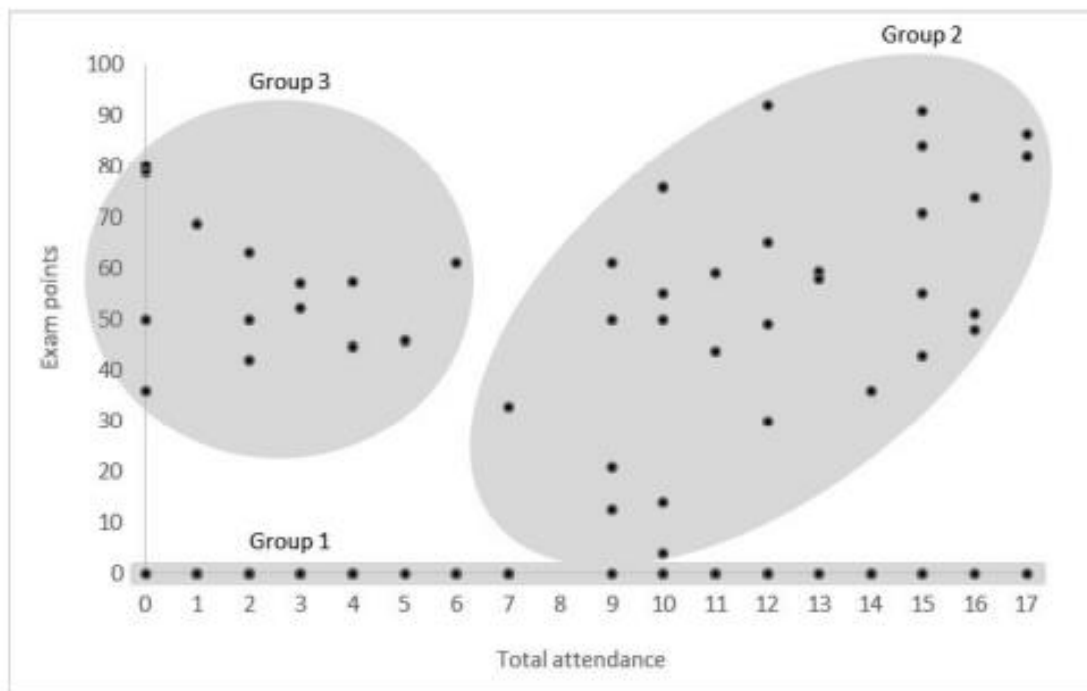
Kučakin, Juričić ja Đambićin (2018) mukaan nykyiset trendit koneoppimisen hyödyntämisessä oppilaitoksen toiminnassa keskittyvät:

1. Opiskelijan tulosten arviointiin.
2. Opiskelijan säilyttämiseen oppilaitoksessa.
3. Opiskelijan tehokkuuden ennustamiseen.
4. Opiskelijoiden testaamiseen.

Koneoppimisen avulla voidaan tarjota apua opettajille ja henkilökunnalle opiskelijan tulosten arvioinnissa, sillä sen avulla voidaan luokitella myös opiskelijan käsin kirjoittamia arviointipapereita. Koneoppiminen pystyy tunnistamaan opiskelijat, jotka mahdollisesti keskeyttävät opintonsa, jolloin oppilaitokset pystyvät tarjoamaan apua näille opiskelijoille. Tutkimuksessa todetaan, että opiskelijoiden keskeyttämisellä voi olla suuret vaikutukset oppilaitoksen suosioon, rahoitukseen ja tilastosijoitukseen.

Lukkarisen, Koivukankaan ja Seppälän (2016) tekemässä tutkimuksessa tutkittiin yliopisto-opiskelijoiden suoriutumista opinnoista suhteessa heidän poissaoloihinsa. Tutkimuksessa hyödynnettiin klusterointi- ja regressiomenetelmiä, joiden avulla aineisto jakaantui kolmeen pääryhmään (kuvio 8):

- 1) Opiskelijat, jotka lopettavat kurssin ennen päättökoetta (group 1)
- 2) Opiskelijat, jotka osallistuvat kurssille ja päättökokeeseen (group 2)
- 3) Opiskelijat, jotka suorittavat kurssin itsenäisesti (group 3).



KUVIO 8 Opiskelijoiden poissaolojen ja kurssimenestyksen vertailu (Lukkarinen ym., 2016).

Tutkimuksessa oli mukana myös muita taustatietoja, kuten esimerkiksi opiskelijoiden ikä, sukupuoli, bonustehtävien pisteet ja tieto siitä, oliko opiskelija suorittanut kurssin esitietona suositellun kurssin. Kuitenkin näistä taustatiedoista opiskelijan poissaolot oli merkittävin tekijä selittämään loppukokeen menestystä. Tutkimuksen tuloksia voidaan hyödyntää kurssien suunnittelussa niin, että opiskelijat saataisiin motivoitua osallistumaan tunneille. Lisäksi kurseja voidaan suunnitella ottamaan huomioon itsenäisesti kurssin suorittavat opiskelijat, jotka eivät osallistu tunneille.

Kučakin ja kollegoiden (2018) mukaan opiskelijan tehokkuuden ennustamisella voidaan löytää opiskelijakohtaisia heikkouksia tietyllä osa-alueella ja tarjota niihin kehitysehdotuksia. Tutkimuksessa löydettiin heikoiten suoriutuvat opiskelijat heidän muiden ja nykyisen kurssin suoritusten perusteella. Nämä opiskelijat huomioitiin opetuksessa eri tavalla, sillä oli todennäköistä, että he saattaisivat epäonnistua. Myös Ciolacun, Tehranin, Beerin ja Poppin (2017) tutkimuksessa koneoppimisen avulla voitiin kategorisoida opiskelijat heidän luontaisen oppimistyylinsä perusteella.

Romeron ja Venturan (2007) mukaan tiedonlouhintaa voidaan hyödyntää tarjoamaan tietoa kurssin ohjaajalle. Palautejärjestelmien avulla voidaan kerätä aineistoa, jotka voidaan yhdistää muihin tietoihin. Tämän tiedon perusteella voidaan esittää kehitysehdotuksia esimerkiksi siitä, kuinka opiskelijoiden oppimista voidaan parantaa tai kuinka kurssin materiaalit voidaan koota paremmin oppimisympäristössä. Järjestelmien tapahtumatietojen perusteella opettajalle voidaan myös esittää tietoa, kuinka opiskelijat käyttäytyvät oppimisympäristössä.

Soinisen (2021) tutkimuksessa ammatillisen oppilaitoksen henkilöstökysely on toteutettu tietovarastoa hyödyntämällä. Henkilöstökyselyn lähtötiedot on haettu tietovarastosta, jossa on mahdollista yhdistää vastaajiin erilaisia tietoja useista lähdejärjestelmistä. Tulosten pohjalta on voitu tehdä syyseuraussuhteiden analysointia Bayesian-luokittelumenetelmällä. Tämän avulla on voitu löytää erilaisia toimenpiteitä koulutuksen järjestäjän henkilöstön toiminnan kehittämiseksi.

Markelloun, Mousouroulin, Spirosin ja Tsakalidiksen (2005) mukaan tiedonlouhinnan avulla opiskelijoille voidaan tarjota adaptiivisempi verkko-opiskeluympäristö. Perinteisessä verkko-oppimisen ympäristössä materiaalit ovat esillä esimerkiksi linkkeinä, jotka kurssin ylläpitäjän suunnittelemassa järjestyksessä ja rakenteessa. Adaptiivisessa järjestelmässä aineisto voidaan personoida opiskelijakohtaisesti heidän tarpeidensa mukaisesti. Adaptiivinen verkko-oppimisympäristö voidaan rakentaa hyödyntämällä tiedonlouhintaa, käyttäjän mallintamista ja profilointia sekä tekoälyä.

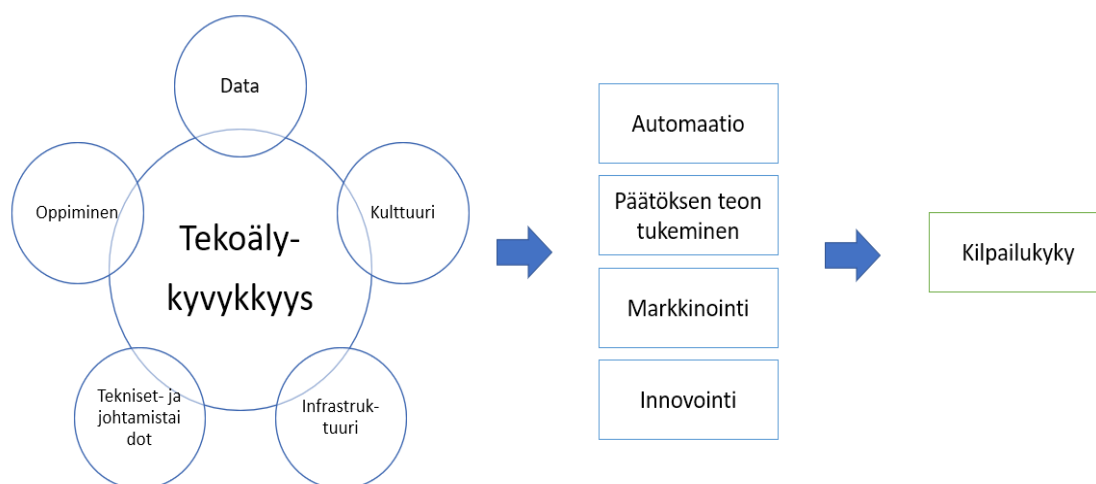
Opetus- ja kulttuuriministeriön oppimisanalytiikan viitekehyksessä (Opetus- ja kulttuuriministeriö, 2021) todetaan, että oppimisanalytiikan käyttäminen on organisaation päätös, ja sen pohjalta organisaatiolle asetetaan myös laillisia vastuita, joihin vaikuttaa käytettävän oppimisanalytiikan muoto. Oppimisanalytiikassa henkilötietolainsäädännön mukaisesti rekisterinpitäjäksi määritellään opetuksen järjestäjä, jonka vastuulla on varmistaa esimerkiksi tie-

tosuojaan liittyvien oikeudellisten kysymysten huomioiminen. Henkilötietojen käsittelyn tulisi perustua rekisterin pitäjän lakisääteisen veloitteen noudattamiseen, julkisen vallan käyttöön tai yleistä etua koskevan tehtävän tekemiseen. Henkilötietojen käsittely voi perustua suostumukseen esimerkiksi lisäpalveluita tarjottaessa.

Oppimisanalytiikan viitekehyksessä todetaan, että oppimisanalytiikan tuloksia tuottavat algoritmit ovat ihmisten rakentamia, ja niiden käyttötarkoitus ja merkitys pitää olla tiedossa. Opetustoimija on vastuullinen analytiikan hyödyntämisestä sekä määrittelee omat velvollisuutensa toimia analytiikan pohjalta. Lain pohjalta voidaan kuitenkin velvoittaa opetustoimijaa tekemään väliintulo esimerkiksi tilanteessa, jossa analytiikka tuottaa tiedon oppijan tuen tarpeesta. Tällöin opetuksen järjestäjällä tai opettajalla voi olla velvoite tarjota oppimisen tukea. Opetuksen järjestäjän tulisi kuvata omat prosessit analytiikan hyödyntämisestä. (Opetus- ja kulttuuriministeriö, 2021).

3 Tutkimuksen teoreettinen viitekehys

Mikalefin ja kollegoiden (2019) teoreettinen viitekehys tekoälykyvykkyudesta ja kilpailukyvyvystä kuvaa ne ulottuvuudet, mitä organisaatiolta vaaditaan tekoälykyvykkyuden saavuttamiseksi, sekä hyödyt, joita tekoälyn sovelluksilla voidaan saavuttaa (kuvio 9). Viitekehyksessä tekoälykyvykkyys jakaantuu viiteen osaan: dataan, tekoälykulttuuriin, infrastruktuuriin, teknisiin- ja johtamistaitoihin sekä oppimiseen. Viitekehymen mukaan organisaatioiden tulisi kehittää ja edistää tekoälykyvykkyyttä jos se haluaa saavuttaa mitään merkittäviä hyötyjä tekoälyn tehtävistä investoinneista.



KUVIO 9 Tekoälykyvykkyuden ja kilpailukyvyvyn viitekehitys (Mikalef ym., 2019).

Tekoälykyvykkyuden ulottuvuudet jakautuvat konkreettisiin (data ja infrastruktuuri), epäkonkreettisiin (tekoälykulttuuri) sekä ihmisten osaamiseen ja tietämykseen. Tekoälykyvykkyuden avulla voidaan saavuttaa hyötyjä esimerkiksi automaation, päätöksen teon tukemisen, markkinoinnin tai

innovoinnin näkökulmasta. Näillä hyödyillä on taas viitekehyksen mukaan yhteys yrityksen tai organisaation kilpailukykyyn tai -etuun.

3.1 Data ja infrastruktuuri

Viitekehyksen mukaan data, eli aineisto, on tärkeässä roolissa tekoälyn toteutuksen kannalta. Sen tueksi tarvitaan myös sopivat teknologiat tiedon käsittelyyn ja säilömiseen. Tiedon laatua korostetaan tärkeimmäksi tekijäksi tekoälyn hyödyntämisessä.

Tietoa voidaan hyödyntää yhdestä tai useammasta järjestelmästä, sekä lisäksi eri järjestelmien tuottaman tiedon hyödyntäminen voi vaatia tiedon puhdistusta ja käsittelyä paremman tiedon laadun saavuttamiseksi. Tämä vaatii sopivan teknologian hyödyntämisestä, jonka avulla tietoa voidaan hallita.

Tiedon käsittely ja analysointi vaatii myös siihen sopivaa infrastruktuuria. Infrastruktuuri tarkoittaa tässä kontekstissa esimerkiksi pilviteknologian hyödyntämistä tiedon prosessointiin sekä riittävän laskentatehon prosessoreita. Tämän lisäksi vaaditaan sopivia tietoliikenneyhteyksiä ja -verkkoja, jotka tukevat tehokkuutta ja skaalautuvuutta.

3.2 Tekniset- ja johtamistaidot

Tekoälykyvykkyyden saavuttamiseksi tarvitaan lisäksi ihmisten tietämystä ja taitoa, joka voidaan jakaa teknisiin- ja johtamistaitoihin. Tekoälyn hyödyntäminen liiketoiminnassa vaatii ihmisiltä kaukokatseisuutta sekä johtamistaitoa. Tällöin tekoälyn sovelluksia voidaan hyödyntää ratkaisemaan liiketoiminnallisia ongelmia. Teknisillä taidoilla tarkoitetaan esimerkiksi osaamista kehittää tekoälymallinnuksen sovellus.

Tekoälyn sovelluksen kehittämiseksi viitekehysessä esitellään kolme pääprofiilia: kouluttajat (*trainers*), selittäjät (*explainers*) ja ylläpitäjät (*sustainers*). Kouluttajat opettajat tekoälyn sovellukselle kuinka sen tulisi toimia, selittäjät asettuvat teknologistien ja liiketoimintajohtajien väliin tarjoamaan selkeyttä ja ylläpitäjien tehtävänä on varmistaa, että tekoälyn sovellukset toimivat kuten pitääkin.

3.3 Organisaation oppiminen ja tekoälykulttuuri

Viimeisenä osa-alueena viitekehysessä esitetään tekoälykulttuuri, joka on tutkimusten mukaan tärkeimpiä tekijöitä tekoälyn hyödyntämisen onnistumisessa. Yrityksen toimintakulttuurin tulisi olla myönteinen tekoälyn sovelluksille ja niiden käyttöönololle.

Viitekehyksen teoria esittää, että yksi päätekijä epäonnistuneelle tekoälyn käyttöönnotolle liittyy tekoälyn kulttuurilliseen vastustamiseen. Myös organisaation muut kilpailevat investointien painopisteet voivat hankaloittaa tekoälyn käyttöönottoa. Tulokset osoittavat, että tekoälykulttuurin omaksuminen ja sitä tukevan strategian kehittäminen ovat kriittisiä tekoälyn hyödyntämiseksi.

3.4 Tekoälyn hyödyntäminen

Viitekehyyksessä esitetään neljä eri osa-aluetta, joihin tekoälyn avulla voidaan tuoda liiketoimintaan lisäarvoa:

1. Automaatio

Tehtäviä ja toimenpiteitä voidaan toteuttaa automaattisesti tekoälyn sovellusten toimesta, joita voi olla esimerkiksi automaattinen käsittely tai keskustelubotit. Näin työntekijöille jää aikaa muuhun ajatustyöhön.

2. Päätöksen teon tukeminen

Päätöksen tekoa voidaan tukea esimerkiksi tekoälyn analyysillä tai neuvoilla.

3. Markkinointi

Tekoälyn avulla voidaan kehittää tarkempia menetelmiä lähestyä asiakasta. Lähestymistapa ja -viesti perustuu hienojakoiseen informaatioon, jota asiakkaasta saadaan.

4. Innovointi

Tekoälyn sovellus voi antaa sisältöä esimerkiksi suunnittelijoille uusien ideoiden kehitykseen.

Tekoälyn tuomalla lisäarvolla näillä osa-alueilla katsotaan olevan vaikutusta organisaation kilpailukykyyn. Tekoälyn hyödyntäminen ja hankkiminen voidaan nähdä organisaatiossa investointina, jolla odotetaan olevan liiketoimintaa parantavia vaikutuksia.

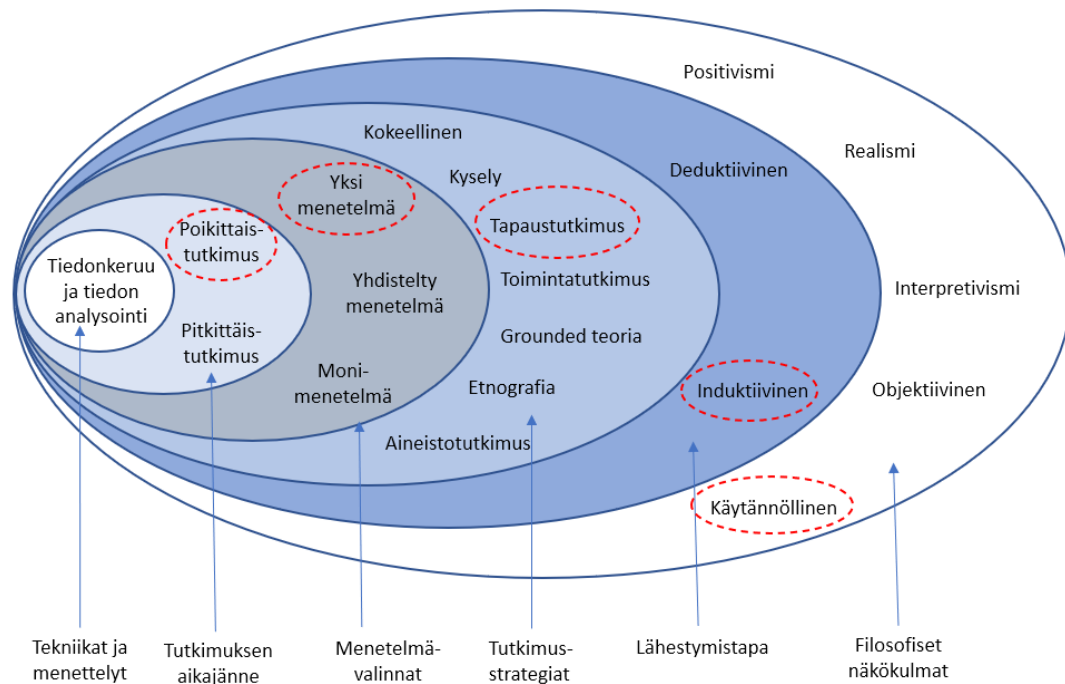
4 Tutkimusmenetelmät

Tässä kappaleessa on kuvattu tutkimuksen toteutuksen tutkimusmenetelmät. Tutkimus on toteutettu tapaustutkimuksena tekemällä havainnot ammatillisen koulutuksen järjestäjän toimintaympäristöstä. Ensimmäisessä kappaleessa esitellään valittu tutkimusstrategia, jonka jälkeen kuvataan kohdeorganisaatio. Tämän jälkeen kuvataan, kuinka tutkimuksessa on tarkoitus kerätä aineistoa ja analysoida sitä.

4.1 Tutkimusstrategia

Tutkimuksen toteutusta voidaan kuvata Saundersin, Lewisin ja Thornhillin (2009) sipulimallin avulla. Sipulimallin avulla voidaan havainnollistaa tutkimuksessa käytettävät menetelmät kerroksittain (kuvio 10). Sipulimallia voidaan purkaa keskeltä ulospäin, jolloin kerrokset ovat

- Tekniikat ja menettelyt
- Tutkimuksen aikajänne
- Menetelmävalinnat
- Tutkimusstrategiat
- Lähestymistapa
- Filosofiset näkökulmat.



KUVIO 10 Sipulimalli (Saunders ym., 2009).

Keskiössä on tiedonkeruu ja tiedon analysointi (tekniikat ja menettelyt), jonka ulompi kerros tarkoittaa kerätyn aineiston aikajänteen. Aineiston kerääminen jakaantuu joko pitkittäis- (aineistoa kerätään useamman kerran) tai poikittäistutkimukseen (aineiston kerääminen on kertaluontoinen). Menetelmävalintaa tarkoittaa, onko tutkimus toteutettu yhdellä menetelmällä, yhdistelmämenetelmänä vai monimenetelmänä. Yksittäinen menetelmä jakaantuu laadulliseen ja määrälliseen tutkimukseen. Yhdistelty menetelmä on yhdistelmää laadullisesta sekä määrällisestä tutkimuksesta. Monimenetelmässä voidaan esimerkiksi luokitella laadullinen aineisto määrälliseen tutkimukseen sopivaksi.

Hirsjärven, Remeksen ja Sajavaaran (2015) mukaan laadullisen ja määrällisen tutkimusmenetelmän eroja voidaan havainnollistaa alla olevan taulukon esittämällä tavalla (taulukko 4).

TAULUKKO 4 Määrällisen ja laadullisen tutkimusmenetelmän erot (Hirsjärvi ym., 2015).

Näkökulma	Määrällinen menetelmä	Laadullinen menetelmä
Tutkijan ja tutkittavan suhde	Etäinen	Läheinen
Tutkimusstrategia	Rakenteellinen	Rakenteeton
Aineiston luonne	Kova, luotettava	Rikas, syvä
Teoria ja tutkimuksen luonne	Teoriaa varmistava	Teoriaa luova

Sipulimallin ulommat kerrokset tarkentavat käytettävän tutkimusstrategian, tutkimuksen lähestymistavan sekä filosofiset näkökulmat. Deduktiivisessa lähestymistavassa siirrytään isommista kokonaisuuksista pienempiin, joka tarkoittaisi, että käytössä olisi jo tiedossa oleva teoria tai malli. Induktiivisessa lähestymistavassa lähtökohtana on yksittäiset havainnot, joista muodostetaan yleistys tai teoria.

Tässä tutkimuksessa aineiston kerääminen on kertaluontoista, joten kyseessä on poikittaistutkimus. Kerättävä ja analysoitava aineisto on luonteeltaan laadullista ja tutkimus on tarkoitus toteuttaa kohde organisaation tapaustutkimuksena. Tämän tutkimuksen lähestymistapaa voidaan pitää induktiivisena, sillä tarkoituksena on edetä yksittäisistä havainnoista yleisempiin väitteeseen (Hirsjärvi ym., 2015). Filosofiselta näkökulmalta tätä tutkimusta voidaan pitää pragmaattisena eli käytännöllisenä, mikä sopii hyvin valittuun tutkimusstrategiaan, sillä se korostaa tutkimuksen tiedon käytännöllistä luonnetta.

Yinin (2009) mukaan tapaustutkimus antaa tutkijoille mahdollisuuden tutkia ja parantaa tapauksen ymmärtämistä sen todellisessa ympäristössä. Tapaustutkimusmenetelmä sopii esillä olevaan tutkimukseen, koska sen avulla voidaan tutkia, kuinka jatkuvia parannusmenetelmiä käytetään ja miten ne näkyvät organisaation toiminnassa. Tapaustutkimuksen yleinen määrittely on vaikeaa eikä sitä voida helposti rajata muihin asetelmiin.

Tapaustutkimuksessa (*case study research*) tutkitaan jotain rajattua kokonaisuutta tai tapahtumaa. Laineen, Bambergin ja Jokisen (2007) mukaan tapaustutkimus luokitellaan tutkimusstrategiaksi tai -tavaksi, eikä tutkimusmenetelmäksi. Tapaustutkimuksessa pyritään tutkimaan, kuvaamaan ja selittämään kohdetta miten- ja miksi kysymysten avulla. Tapaustutkimuksessa käytetään erilaisia menetelmiä tiedon keräämiseen ja analysointiin, joten se ei ole pelkästään aineistonkeruun tekniikka. Tapaustutkimus sopii monelle eri tieteenalalle (Yin, 2009).

Tapaustutkimuksen tavoitteena on tutkimuskohteen piirteiden systemaattinen, tarkka ja totuudenmukainen kuvailu (Anttila, 1996). Hirsjärven ja kollegoiden (2015) mukaan tutkimusstrategiat jakaantuvat kolmeen eri tyyppiin:

- Kokeellinen tutkimus
- Kvantitaavinen survey-tutkimus
- Kvalitatiivinen tapaustutkimus

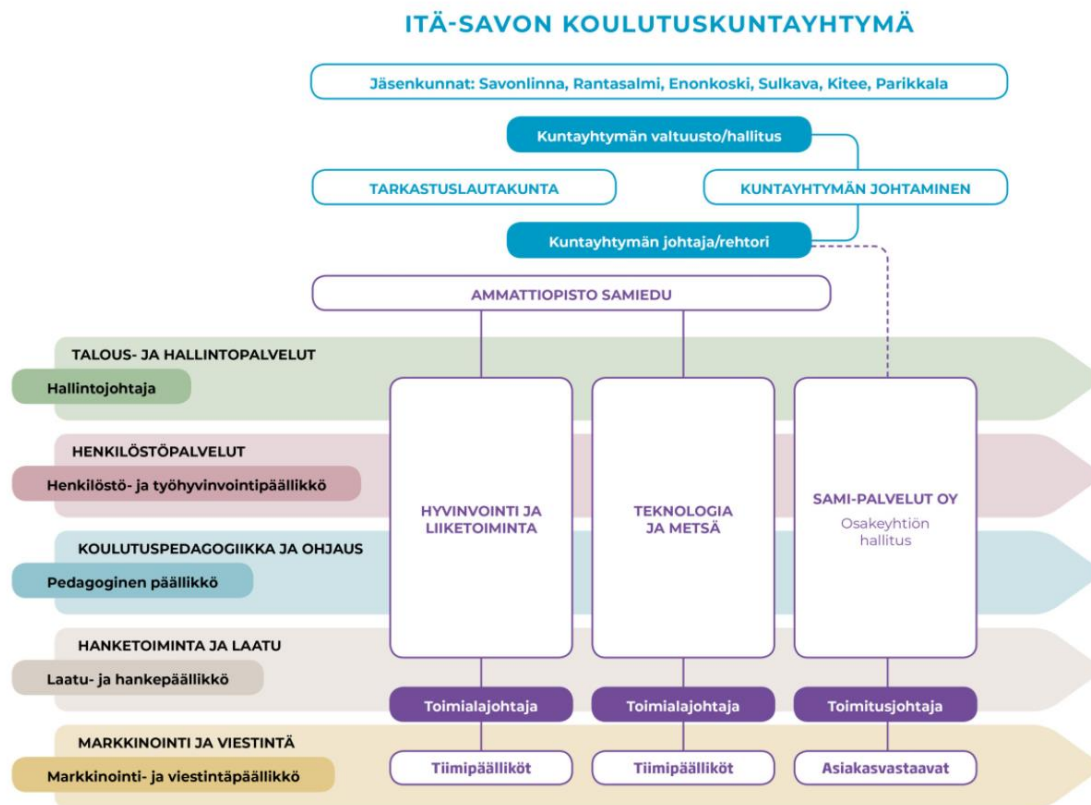
Tapaustutkimus koostuu empiirisistä havainnoista, jolla tutkitaan ilmiötä ajan, paikan tai jonkin muun kriteerin mukaan (Creswell, 1998).

Tässä tutkimuksessa kiinnostuksen kohde on reformin aiheuttamat muutokset koulutuksen järjestäjän toiminnassa, sekä nykyaikaisen teknologian antamat mahdollisuudet opetustoiminnan kehittämisessä. Darken, Shanksin ja Broadbentin (1998) mukaan tapaustutkimusta voidaan käyttää tietojärjestelmien tutkimisessa organisaatioissa. Tapaustutkimuksella voidaan ymmärtää tietojärjestelmää syvällisemmin sen kontekstissa, mutta sen heikkoutena on kui-

tenkin heikko yleistettävyys sekä aineiston keräämisen ja analysoinnin subjektiivisuus. Aineistoa voidaan kerätä haastatteluilla, kyselyillä, havainnoimalla ja valmiilla aineistoilla.

4.2 Case-kuvaus

Tutkimuksen kohdeorganisaatio on Itä-Savon koulutuskuntayhtymän ylläpitämä ammattiopisto Samiedu (kuvio 11). Ammattiopisto Samiedu sijaitsee Savonlinnassa, mutta heidän opintotarjontaansa kuuluu myös verkossa tapahtuva etäopiskelu. Samiedulla opiskelee vuosittain noin 1 600 tutkinto-opiskelijaa ja kokonaisasiakasmäärä on lähes 4 500 henkilöä. Samiedun visiona on olla haluttu alueellinen ja kansainvälinen huippukouluttaja ja yhteistyökumppaneille kumppanuusverkoston rohkea uudistaja (Samiedu, 2021).



KUVIO 11 Itä-Savon koulutuskuntayhtymän organisaatorakenne (Samiedu, 2021).

Samiedun toiminta-ajatuksena on tukea opiskelijan ammatillista, yleisivistävää ja persoonallisuuden kasvua, kouluttaa ammattitaitoista henkilöstöä työelämän muuttuviin tarpeisiin ja jatko-opintoihin, turvata koulutuksen monipuolinen alueellinen saatavuus, kehittää työelämää ja yrittäjyyttä sekä vastata niiden osaamistarpeisiin, edistää kansainvälistymistä ja edistää kestävästä kehi-

tyksen periaatteita. Näitä tehtäviä toteuttaakseen Itä-Savon koulutuskuntayhtymä ylläpitää Ammattiopisto Samiedua ja SAMI-Palvelut Oy:tä sekä muita koulutustehtävän kannalta tarpeellisia toimintoja ja kiinteistöjä. Samiedun arvot ovat (Samiedu, 2021):

- Välittävä ja oppimiseen kannustava vuorovaikutus
- Ohjaava ja valmentava oppimis- ja työilmapiiri
- Uudistuva ja osallistava toimintakulttuuri
- Luotettava ja vastuullinen yhteistyökumppani

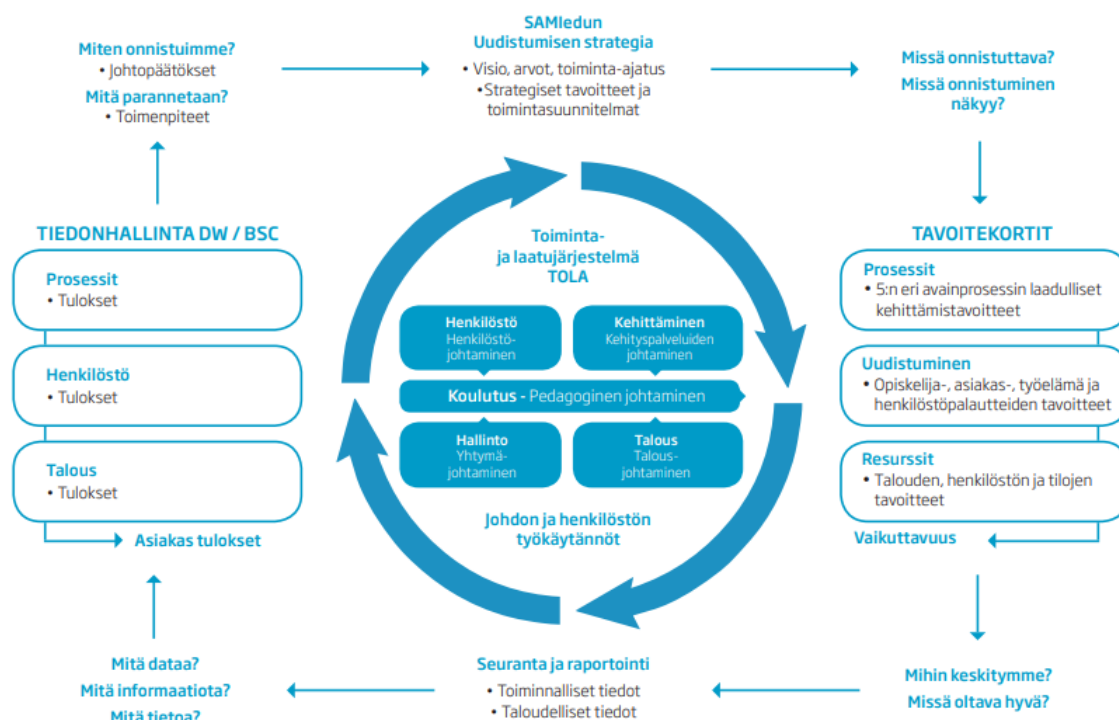
Samiedun strategiset päätavoitteet ja menestystekijät vuosille 2018–2021 on jaettu koulutuksenjärjestäjän, organisaation, verkostojen ja palveluiden tasoille (kuvio 12). Koulutuspalveluiden päätavoite on opiskelijan työllistymiseen tai jatko-opintoihin tähtäävä henkilökohtaisesti luotu opintopolku. Strategialla vastataan ammatillisen koulutuksen reformin ja toimintaympäristön muuttuviin haasteisiin. Strategiassa todetaan, että uuden toiminta- ja rahoituslain tuoma muutos edellyttää koulutuksen järjestäjältä rakenteiden ja prosessien rohkeaa uudistamista. Rahoituslain muutoksen johdosta koulutuksen resurssit määräytyvät jatkossa erilaisten tuloksellisuustekijöiden seurauksena, kuten opiskelijoiden suoritusten (suoritusrahoitus) ja opiskelijoiden ja työelämän antamasta palautteesta koulutuksen laadusta (vaikuttavuusrahoitus).

KOULUTUKSEN JÄRJESTÄJÄ	ORGANISAATIO	KOULUTUSPALVELUT
<p>Strateginen päätavoite</p> <p>Monialainen ja monipuolisesti koulutusta tarjoava järjestäjä, joka pystyy laajasti vastaamaan työelämän muuttuviin osaamistarpeisiin sekä toimintaympäristön muuttuviin palvelutarpeisiin.</p> <p>Talouden tasapainottaminen.</p> <p>Menestystekijät</p> <ul style="list-style-type: none"> • Hyvin toimiva omistajaohjaus. • Laaja-alainen koulutustarjonta ja alueellinen koulutuksen saatavuus. • Kustannusrakenteen joustavuus rahoituspohjaa ja koulutustarpeita vastaavaksi. • Ammattitaitoa vahvistavat tutkimuksen osat 	<p>Strateginen päätavoite</p> <p>Organisaation osaamisala rakenteen ja sitä tukevien yhtymäpalveluiden prosessien uudistaminen tukemaan seuraavia keskeisiä päämääriä:</p> <ul style="list-style-type: none"> • Asiakaslähtöisyys. • Laadunhallinta. • Vaikuttavuus. <p>Menestystekijät</p> <ul style="list-style-type: none"> • Jatkuvasti kehittyvä johtaminen ja asiantuntija organisaatio ylläpitävät ammatillisia, toiminnallisia ja taloudellisia valmiuksia riittävän kilpailukyvyyn varmistamiseksi. 	<p>Strateginen päätavoite</p> <p>Työllistymiseen ja/ tai jatko-opintoihin tähtäävä työelämän ja muiden koulutuksen järjestäjien kanssa yhteistyössä henkilökohtaistettu opintopolku.</p> <p>Hyvin toimivat ja integroidut työpaikka-, oppilaitos ja verkko-oppimisen oppimisympäristöt.</p> <p>Menestystekijät</p> <ul style="list-style-type: none"> • Uudistuva ja vetovoimainen koulutus- ja palvelutarjonta. • Varmistetaan opiskelijoiden riittävä ammatillinen osaaminen. • Ympärivuotinen oppiminen/opintosuoritusten ympärivuotinen kertymä. • Vahvistetaan työpaikalla tapahtuvan oppimisen ohjausta.
HENKILÖSTÖPALVELUT	KEHITTÄMISPALVELUT	VERKOSTOT
<p>Strateginen päätavoite</p> <p>Muutosvalmennuksella saavutettu opetushenkilöstön pedagoginen ja didaktinen asiantuntemus ja koko henkilöstön laaja-alainen ammatti-identiteetti.</p> <p>Työhyvinvointia edistävän työn muutokseen tarvittavien ammatillisten valmiuksien parantaminen.</p> <p>Menestystekijät</p> <ul style="list-style-type: none"> • Sujuvan työn esteiden poistaminen tiimityössä. • Vuosityöaika. • Hyvin toimiva esimiestyö ja prosessien johtaminen. • Työhyvinvointia työtä kehittämällä. • Henkilöstön sitoutuminen, esimiestyön ja työhyvinvoinnin arviointi. 	<p>Strateginen päätavoite</p> <p>Strategiaan perustuvien kehittämissuunnitelmien ohjaus ja seuranta.</p> <p>Prosessien jatkuvan laadun parantamisen toimivuuden varmistaminen ja ohjaus.</p> <p>Laadunvarmistukseen tarvittavien työprosessien ylläpitäminen ja ohjaus.</p> <p>Menestystekijät</p> <ul style="list-style-type: none"> • Jatkuvan laadun parantamisen työkäytäntöjen kehittäminen ja arviointi. • Osaamisalakohtaisen toiminnan tulosten pohjalta tapahtuvan kehittämisen ohjaus ja seuranta. 	<p>Strateginen päätavoite</p> <p>Roolin vahvistaminen seudullisissa verkostoissa.</p> <p>Vahvat kansalliset ja kansainväliset kumppanuudet.</p> <p>SAMledun brändin vahvistaminen kansallisena ja kansainvälisenä kehittäjänä ja koulutuksen järjestäjänä.</p> <p>Menestystekijät</p> <ul style="list-style-type: none"> • Seudullisen ja kansallisen yhteistyön korostaminen. • Arvostettu kumppanuus kehittämissuunnitelmien verkostoissa. • Kansainväliset kumppanuudet koulutuspalveluiden edellytysten turvaamiseksi.

KUVIO 12 Samiedun strategiset päätavoitteet ja menestystekijät 2018–2021 (Samiedu, 2021).

Samiedu toimeenpanee uudistumisen strategiaa kuntayhtymän ja ammattopiston vuosisuunnittelussa, jossa strategiset linjaukset siirretään organisaation eri tasoille. Strategisilla tavoitekorteilla määritellään tuloksellisuuden arvioinnin perusteet ja indikaattorit. Strategian ja jatkuvan laadunparantamisen prosessit ja työkäytännöt on kuvattu kaaviossa (kuvio 13).

STRATEGIAN JA JATKUVAN LAADUNPARANTAMISEN PROSESSIT JA TYÖKÄYTÄNNÖT



KUVIO 13 Samiedun uudistumisen strategian (2018-2021) ja jatkuvan laadunparantamisen prosessit ja työkäytännöt (Samiedu, 2021).

Samiedu käyttää strategian viitekehyksinä BSC-tavoitekortteja (*Balanced Scorecard*), jatkuvan laadunparantamisen toimintamalleja, IMS-toimintajärjestelmää sekä DW-tiedonhallintaa. Demingin laatuympyrä PDSA (Plan-Do-Study-Act) on systemaattinen tapa kuvata tuotteen, prosessin tai palvelun jatkuvaa parantamista. IMS-toimintajärjestelmä on ohjelmisto, joka kokoaa organisaation prosessikuvaukset, työnohjauksen dokumentaatiot, käsikirjat ym. DW-tiedonhallinnalla tarkoitetaan Samiedun tietovarastoa.

4.3 Aineiston käsittely

Tämän tutkimuksen aineiston kerääminen koostuu valmiista aineistoista ja dokumenteista (Hirsjärvi ym., 2015). Valmiilla aineistoilla tarkoitetaan tässä tapauksessa esimerkiksi kohdeympäristössä olevia teknisiä sovelluksia, tietokantoja ja muita ohjelmistoprojekteja. Lisäksi aineiston keruussa käytetään erilaisia dokumentaatioita sekä niihin liittyviä asiakirjoja. Dokumentaatiot voivat olla esimerkiksi järjestelmien teknisiä dokumentaatioita tai niihin liitettyjä muistioita, joita voidaan käyttää tätä tutkimusta tehdessä. Kaikki tutkimuksessa käytettävä aineisto ei välttämättä ole vapaasti julkaistavissa, kuten järjestelmien tekniset dokumentaatio ja sopimusasiakirjat.

Tieto kerätään siinä muodossa kuin se on saatavilla, eikä sitä täydennetä esimerkiksi haastatteluilla tai kyselyillä. Aineiston kerääminen ja analysointi rajataan tapausorganisaation IT-ympäristöön, joka liittyy opiskelija tietojen keräämiseen, raportointiin ja analysointiin. Tutkijalla on pääsy kohteen palvelinympäristöön. Tutkimuksessa ei käydä läpi muita organisaation toimintoja ja ympäristöjä, kuten markkinointia tai myyntiä. Aineistoa voidaan kuitenkin verrata yleisellä tasolla organisaation strategiaan.

4.4 Aineiston analysointi

Tutkimuksen teoriaosuudessa on luotu käsitys siitä, minkälaisia työkaluja, menetelmiä ja toimintamalleja on olemassa tekoälyn sovellusten toteuttamiseksi ja hyödyntämiseksi. Aineiston analysoinnissa hyödynnetään Mikalefin ja kollegoiden (2019) tekoälykyvykkyyden ja kilpailukyvyyn viitekehysten esittämiä ulottuvuuksia, joiden avulla organisaatio pystyy saavuttamaan tekoälykyvykkyyden. Lisäksi viitekehyksessä esitetään myös hyödyt, joita tekoälyllä voidaan saavuttaa, joita verrataan kohdeorganisaatiosta löytyneisiin tuloksiin.

Tutkimus lähtee liikkeelle organisaation tiedonhallinnasta ja etenee siitä sen hyödyntämiseen raportoinnissa, analysoinnissa ja tekoälymallinnuksessa. Tutkimuksessa pyritään analysoimaan myös päätökset ja toimenpiteet, mitkä ovat johtaneet kyseisten mallinnusten ja työkalujen käyttöön. Aineistoa pyritään myös analysoimaan siinä hetkessä, kun se on tuotettu. Täten sitä voidaan tutkia retrospektiivisesti muuttuneeseen toimintaympäristöön.

Kerätyn aineiston pohjalta tehdään tuloksina primaareja empiirisiä kontribuutioita (*primary empirical contribution, PEC*), jotka tiivistävät tehtyjä empiirisiä havaintoja. Näitä havaintoja käsitellään tutkimuksen johtopäätöksissä.

5 Tulokset

Tässä luvussa käydään läpi tutkimuksen tulokset. Alustavana tutkimustuloksena voidaan dokumentaatioiden perusteella todeta, että kohdeorganisaatiossa on käytössä automatisoidut tiedolla johtamisen työkalut, joiden yhtenä osa-alueena on opiskelijoiden tekoälymallinnus. Aineiston perusteella tekoälymallinnus pohjautuu koulutuksen järjestäjän tietovarastossa sijaitsevaan tietoon.

Toimintatavat, rakenteet ja tekniset ratkaisut, joilla tämä kokonaisuus on toteutettu ja hyödynnetty, käydään tässä luvussa läpi. Kappaleet on jaoteltu tutkimusasetelmassa kuvattuun järjestykseen, lähtien organisaation tiedon hallinnasta toteutettuihin sovelluksiin ja niiden hyödyntämiseen. Kappaleiden lopussa esitellään kappaleen tuloksiin liittyviä primäärejä empiirisiä kontribuutioita (PEC).

5.1 Operatiiviset järjestelmät

Kohdeorganisaatiolla on käytössä useita operatiivisia järjestelmiä. Näitä on esimerkiksi oppilashallintoa varten käytössä oleva oppilashallintojärjestelmä Primus, kirjanpitoa ja taloudenhallintaa varten käytössä on Intime ja henkilöstön palvelussuhde- sekä palkkaustiedot ovat Personec -järjestelmässä. Taulukossa on esitetty organisaation käytössä olevat ohjelmistot ja niiden käyttötarkoitus (taulukko 5). Tämä kattaa vain osan käytössä olevien järjestelmien ja palveluiden kokonaisuudesta.

TAULUKKO 5 Samiedun käytössä olevat tietojärjestelmät ja niiden käyttötarkoitus.

Järjestelmä	Käyttötarkoitus	Tietosisältö
Primus	Oppilashallinto	Opiskeluoikeudet, suoritukset, poissaolot ja muut opiskelijamerkinnot
Kurre (Primus)	Tilahallinto	Tilojen varaukset (resursointi), lukujärjestykset
Intime	Taloushallinto	Tositteet,

Personec F	Henkilöstöhallinto	tositteiden kohteet (projektit, kustannuspaikka, toiminto ym.), organisaatorakenne Palvelussuhteet, palkat, poissaolot, työajat ja työehtosopimukset
TCD Survey	Kyselyt ja palautteet	Henkilöstökyselyt, opiskelijakyselyt, ilmoittautumiset ym.
TAIKA	Työajanseuranta	Henkilöstön työajankirjaukset, opettajien vuosityöaika, hanketyöaika, kustannusjakauma

Organisaation eri yksiköillä voi olla käytössä myös erilaisia palveluita ja ohjelmistoja pienemmässä roolissa, mitkä eivät vaikuta koko organisaation tietoarkkitehtuuriin merkittävästi. Myös ulkopuoliset palvelut, kuten käytössä oleva tiedostojen tallentamisen pilvipalvelu tai organisaation intra ja siellä olevat dokumentit eivät ole välttämättä organisaation liiketoimintatiedon hallinnassa merkittävässä roolissa tässä tutkimuksessa.

Opiskelijahallintoon liittyvät opiskelijoiden perustiedot, kuten opinto-oikeus ja erilaiset suoritukset sijaitsevat Primus-järjestelmässä. Kuitenkin tämän lisäksi organisaatiolla on käytössään erilaisia sähköisiä oppimisympäristöjä, kuten Moodle, jonka tietosisältönä on erilaiset kurssikohtaiset suoritukset ja tehtävät. Tutkimuksen toteutuksen aikana Moodle-järjestelmä ei ole vielä organisaation tietoarkkitehtuurin kannalta merkittävässä roolissa, joten se on jätetty taulukon ulkopuolelle.

Talous- ja henkilöstöhallinnon järjestelmät ovat saman toimittajan ohjelmistoja. Näiden ohjelmistojen välillä on palveluntarjoajan puolelta erilaisia ominaisuuksia ja toiminnallisuuksia, joiden avulla nämä järjestelmät pystyvät toimimaan paremmin keskenään. Esimerkiksi kustannuslaskennassa käytettävät tekijät, joita henkilöstön palkanlaskennassa voidaan käyttää ovat samoja, mitä kirjanpitojärjestelmässä on. Toisin sanoen, projektityöntekijän käyttämä kustannuspaikka ja muut kodistustekijät löytyvät myös kirjanpitojärjestelmästä.

Taloushallinnonjärjestelmä pitää sisällään organisaatorakenteen. Se on hierarkkinen rakenne, jonka avulla esimerkiksi tuloja ja kustannuksia voidaan raportoida yksiköittäin. Muut lähdejärjestelmät, kuten oppilashallintojärjestelmä Primus, ei pidä sisällään tätä hierarkiaa. Järjestelmässä ylläpidetään vain yksiköiden alinta tasoa, eli kustannuspaikkaa, johon opiskelijat kohdistuvat.

Kyselyohjelmiston avulla organisaatio voi seurata omien toimintojensa toimivuutta esimerkiksi palautteiden avulla. Palautteet voivat olla niin organisaation asiakkailta (opiskelijoilta) tai omalta henkilökunnalta (esimerkiksi työolobarometrit). Kyselyt voidaan nähdä myös irrallisena osana organisaation tietoarkkitehtuurista, joiden tarkoitus on mitata sen toimivuutta, mutta järjes-

telmän avulla voidaan myös kerätä informaatiota ja kytkeä se osaksi kokonaisuutta. Kohdeorganisaatio käyttää osassa kyselyitään yksilöiviä tunnistetietoja, joiden avulla tiedot voidaan yhdistää muuhun liiketoimintatietoon. Näitä tunnistetietoja on esimerkiksi opiskelijan tai opettajan tunniste, joka tallennetaan kyselyvastaukseen metatietona.

Työajanseurannan avulla organisaatio seuraa henkilöstön työajantoteutumista kustannuspaikka-tasolla. Samiedu on siirtynyt opettajien vuosityöaikaan, jonka tarkoituksena on seurata opetushenkilöstön työaika suhteessa vuodeksi laadittuun vuosityöajansuunnitelmaan. TAIKA työajanseurantaohjelmiston avulla organisaatio seuraa opetushenkilöstön kirjaamaa työaika suhteessa heille laadittuihin suunnitelmiin. Myös työajanseurannassa käytetään samoja kohteita, kuten kustannuspaikkoja ja projekteja, jotka ovat taloushallinnon järjestelmissä.

Kustannuspaikka on kohdeorganisaation alin seurantataso, joka on sama useammassa eri lähdejärjestelmässä. Kustannuspaikkatietoa voidaan tarkentaa lisäksi myös muilla kohteilla, kuten toiminnolla tai projekteilla. Näiden kohteiden avulla organisaatio voi tehdä raportoinnin omasta toiminnastaan niin oman toiminnan seuraamista varten, kuin myös viranomaisraportointiin, joita koulutuksen järjestäjältä velvoitetaan vuosittain tekemään.

Henkilöiden tietoja käytetään useammassa eri järjestelmässä. Jotkin järjestelmät käyttävät henkilöiden yksilöimiseen henkilötunnusta. Esimerkiksi Primuksessa on opettajien tietoja, joita käytetään opiskelijan vastuuhajaajan yksilöimiseen. Opettajien varaukset ovat Kurre-tilahallintajärjestelmässä. Opettajien palvelussuhteisiin liittyvät perustiedot, kuten osa-aikaisuudet ja muut palkkaukseen liittyvät tekijät ovat Personecissa. Opettajien vuosityöajanseuranta tapahtuu vastaavasti työajanseurantaohjelmistossa. Lisäksi opettaja voi kerätä palautetta omasta opetuksestaan Survey-järjestelmällä.

Kerätyn aineiston perusteella voidaan todeta, että tietoarkkitehtuurin näkökulmasta organisaation eri toimintoja varten käytössä olevat ohjelmistot ja palvelut ovat itsenäisiä. Palveluissa on kuitenkin käytössä samoja käsitteellisiä tietoja, kuten kustannuspaikka, mutta ne ovat järjestelmäkohtaisesti hallittuja.

Esitetyn tietoarkkitehtuurin perusteella voidaan todeta, että organisaatiolla on olemassa useita eri operatiivisia ohjelmistoja, joiden tarkoituksena on hyödyntää pääsääntöisesti yhtä käyttäjäryhmää. Näihin järjestelmiin kerääntyä aineistoa niiden käytön myötä. Tämä voidaan tiivistää empiiriseksi havainnoksi:

PEC1: Organisaatiolla on eri käyttötarkoituksia varten useita operatiivisia ohjelmistoja, joihin muodostuu aineistoa.

5.2 Tiedonhallinta

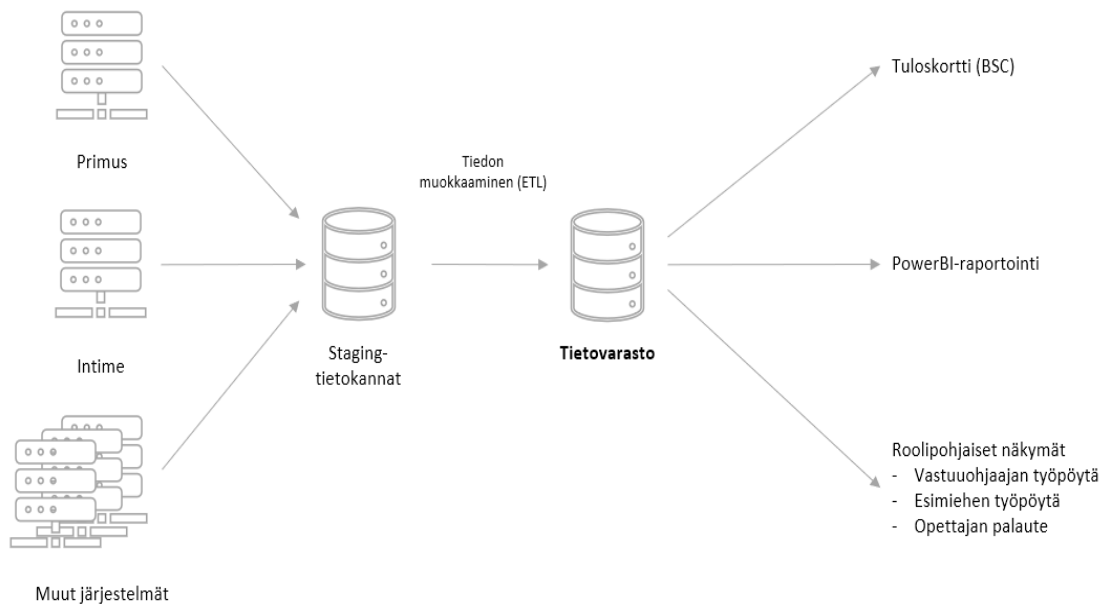
Vuoden 2014 hankintapäätöksen perusteella kohdeorganisaatiossa aloitettiin tietovarastoprojekti, jonka tavoitteena on yhdistellä eri lähdejärjestelmien tiedot

yhtenäiseen tietokantaan (*tietovarastoon*). Tietovaraston tarkoituksena on tässä vaiheessa ollut organisaation eri toimintojen raportoinnin yhtenäistäminen, tiedon laadun parantaminen sekä sen tarjoaminen ajantasaisemmin päätöksen tekoa varten. Lisäksi tavoitteena on ollut päällekkäisten ja manuaalisten työvaiheiden ja tehtävien poistaminen.

Dokumentaation perusteella vuoden 2014 projektissa on toteutettu järjestelmäintegraatiot useasta lähdejärjestelmästä organisaation omaan tietovarastoon. Integraatiossa mukana olleet tekniset järjestelmät ovat siis olleet:

- Primus (opiskelijahallinto)
- Kurre (tilavaraukset, opiskelijoiden ja opettajien kalenterit)
- Intime (taloushallinto)
- Personec (henkilöstöhallinto)
- TCD Survey (kysely / palautejärjestelmä)

Dokumentaation mukaan näistä lähdejärjestelmistä on toteutettu päivittäin tapahtuva tiedonsiirto kohdeorganisaation tietovarastoon. Toteutettu integraatioprosessi on kuvattu alla kaaviona (kuvio 14).



KUVIO 14 Samiedun järjestelmäintegraatiot tietovarastoon.

Lähdejärjestelmien tiedot ladataan joko määrämuotoisena tiedoston-tiedonsiirtona tai ohjelmistorajapintojen kautta erillisinä hakuina lähdejärjestelmästä ja sen kyvykkyydestä riippuen. Dokumentaatiossa on käsitelty lähdejärjestelmien tietojen siivoamista sekä muokkaamista ennen tietovarastoon lataamista, minkä tarkoituksena on parantaa tiedon laatua. Tämä työvaihe on suoritettu integraatioiden ETL-työvaiheessa.

Lähdejärjestelmistä haetut tiedot ovat vaatineet siivoamisen lisäksi myös muokkaamista ETL-työvaiheen aikana. Lähdejärjestelmä Primus esimerkiksi tarjoaa taulukkomuotoista tietoa oman toimintansa mukaisesti erottimella eroteltuna (taulukko 6). Alkaen-kenttä pitää sisällään erotinmerkillä eroteltuna alkamispäivämäärät. Rahoitusmuoto-kenttä pitää vastaavasti sisällään tiedon siitä, mistä rahoitusmuodosta on ollut kyse.

TAULUKKO 6 Primus lähdejärjestelmän tuottamia tietoja taulukkomuodossa (CSV).

Opiskelija	Rahoitus alkaen	Rahoitusmuoto
101	1.1.2019;1.1.2020	R1;R2
102	1.1.2019	R1
103	30.1.2019;18.6.2020	R3;R3

Se, että nämä taulukkomuotoiset kentät saadaan yhdistettyä toisiinsa, vaatii samaisten taulukkoindeksien tutkimista toisesta sarakkeesta, jonka jälkeen tiedot saadaan muutettua tietokantaan ladattavaan muotoon (taulukko 7).

TAULUKKO 7 Esimerkki, kuinka Primus lähdejärjestelmän tiedot on muokattu tietokantaan ladattavaan muotoon.

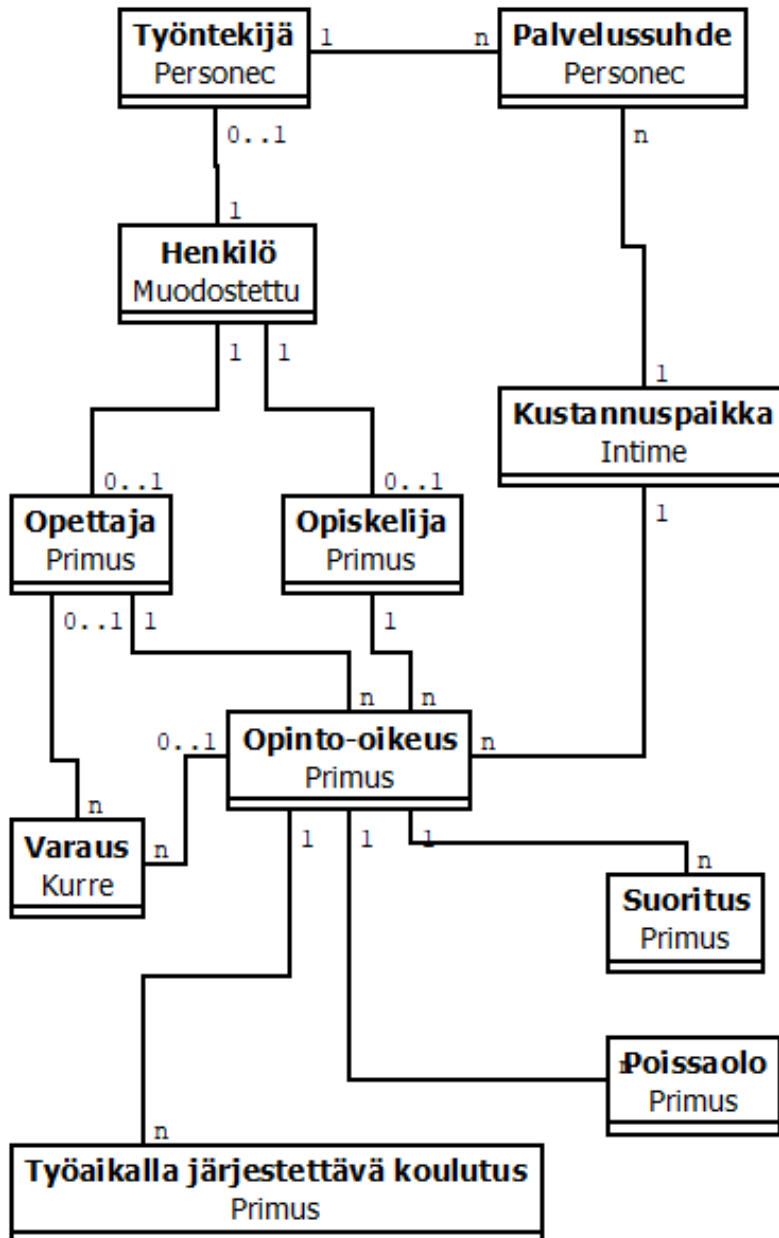
Opiskelija	Rahoitus alkaen	Rahoitusmuoto
101	1.1.2019	R1
101	1.1.2020	R2
102	1.1.2019	R1
103	30.1.2019	R3
103	18.6.2020	R3

Primus-järjestelmässä on vastaavia taulukkomuotoisia kenttiä useammassa eri rekisterissä käytössä, ja ne on purettu tietovaraston latausvaiheessa samalla tavalla.

Tietokannan integraatioissa on myös otettu huomioon kunkin tiedon lähdejärjestelmä. Kustannuspaikka-tietoa pidetään yllä kaikissa lähdejärjestelmissä, mutta tietokantaan lataukset tehdään vain taloushallinnon järjestelmän kautta. Dokumentaatiossa tämän perusteella kustannuspaikkojen nimitykset luetaan siis lähtökohtaisesti taloushallinnon siirtämien tietojen kautta, ja muiden lähdejärjestelmien kustannuspaikoista hyödynnetään vain sen kohdistetekijää (eli kustannuspaikan tunnusta). Näin ollen kustannuspaikan nimen muokkaaminen luetaan vain yhdestä järjestelmästä, eikä esimerkiksi Primukseen lisätyt uudet (tai virheelliset) kustannuspaikat ole raportoinnissa mukana, ennen kuin ne on lisätty taloushallinnon järjestelmään.

Ladatut tiedot on jäsenelty tietovarastoon sen tietomallin mukaisesti. Mallin tarkoituksena on yhdistää useamman eri lähdejärjestelmän sisältämät tiedot sekä pyrkiä normalisoimaan tätä rakennetta tietokannassa. Primus ja henkilöstöhallintojärjestelmä Personec sisältävät molemmat henkilöihin liittyviä tietoja (opettajat ja henkilökunta). Tietomallin avulla nämä tiedot on saatu

yhdistettyä niin, että henkilö on tietokannan rakenteessa vain kerran, jolla voi taas olla roolien kautta yhteyksiä eri järjestelmissä sijaitseviin tietoihin. Kaaviossa (kuvio 15) on kuvattu pelkistetysti tietokannan rakennetta talous-, henkilöstö- ja opiskelijahallintojärjestelmien näkökulmasta.



KUVIO 15 Pelkistetty näkymä tietovaraston rakenteesta.

Kuviosta voidaan todeta, että jokaisen järjestelmän tiedot ovat yhteydessä toisiinsa taulujen välisten relaatioiden kautta, eikä järjestelmien tiedot ole vain irrallisesti tallennettu tietokantaan.

Samiedu on voittanut Opetushallituksen ammatillisen koulutuksen laatu-palkinnon vuosina 2003, 2007, 2011 ja 2017. Laatu-palkinnon painopiste vaihtuu

vuosittain, ja vuoden 2017 laatupalkintokilpailun teema oli ”tiedolla johtaminen” (Opetushallitus, 2021a). Vuoden 2017 arviointiperusteissa (Opetushallitus, 2021b) todetaan muun muassa

- SAMledussa on kehitetty tavoitteellisesti tiedolla johtamisen kokonaisarkkitehtuuria.
- Tietovaraston (DW) avulla saadaan yhdistettyä tulostietoa strategiatyön tueksi.
- Tiedolla johtaminen tukee organisaation tavoitteiden saavuttamista kaikilla organisaation tasoilla. Reaaliaikaisen tiedon saatavuus mahdollistaa prosessien johtamisen.

Laatupalkinnon tarkoitus on kannustaa koulutuksen järjestäjiä toiminnan laadun arviointiin ja jatkuvaan parantamiseen. Laatupalkinnon saanut koulutuksen järjestäjä ei voi hakea palkintoa uudestaan seuraavana kolmena vuotena.

Kappaleen löydösten perusteella voidaan todeta yksi ensisijainen empiirinen havainto (PEC2). Aineiston perusteella tietovarasto on keskeisessä roolissa organisaation tiedonhallinnassa ja sen muokkaamisessa hyödynnettävämpään muotoon. Löydösten perusteella voidaan myös todeta, että operatiivisen järjestelmän rajapinnan kautta haettavaa aineistoa täytyy muokata, ennen kuin sitä voidaan hyödyntää.

PEC2: Tietovaraston avulla voidaan yhdistää useamman lähdejärjestelmän aineistot yhtenäiseen tietokantaan. Integraatioiden avulla voidaan muokata lähdejärjestelmien tuottamaa aineistoa käsiteltävämpään muotoon, esimerkiksi raportointia varten.

5.3 Tekoälymallinnus

Tietovarastopalvelimelle on toteutettu koneoppimisen malli, joka ennustaa opiskelijan todennäköisyyttä valmistua tai keskeyttää opintonsa. Koneoppimisen malli hyödyntää tietovarastoon ladattua tietoa. Dokumentaation perusteella koneoppimisen malli on rakennettu yhteistyössä Jyväskylän yliopiston IT-tiedekunnan sekä TCD Consulting and Research Oy:n kanssa. Samiedu on tehnyt huhtikuussa 2020 päätöksen osallistua tähän kehittämishankkeeseen, jonka tavoite on kuvailtu seuraavasti:

Hankkeen tavoitteena on selvittää opiskelijoiden eroamisen ennustamiseen liittyvät tekijät, jonka perusteella toteutetaan koneoppimisen (machine learning) malli.

Tällä sopimuksella koulutuksen järjestäjä on myös suostunut siihen, että heidän opiskelija-aineistoaan on voitu käyttää hankkeessa tehtävää tutkimusta varten. Opiskelija-aineisto on toimitettu niin, ettei yksittäistä opiskelijaa voida aineistosta jälkikäteen tunnistaa. Konkreettisenä hyötynä hankesuunnitelmassa esitetään, että koulutuksen järjestäjä pystyy ennakoimaan mahdollisesti keskeyttäviin opiskelijoihin ajoissa. Hanketoiminta kuuluu Samiedun organisaatiossa kehittämispalveluihin, jonka strategiaan päätavoitteisiin kuuluu strategiaan perustuvien kehittämisprosessien ohjaus ja seuranta sekä prosessien jatkuvan laadun parantaminen.

Jyväskylän yliopisto on ollut mukana tekoälymallin kehittämisessä tutkimusorganisaationa, jonka tarkoitus on ollut tutkia ja etsiä niitä tekijöitä, jotka vaikuttavat opiskelijan opintomenestykseen. TCD Consulting and Research Oy on tuoteistanut koneoppimisen mallin, sekä asentanut ja ajastanut sen tietovarastopalvelimelle. Kehitetty mallinnus on otettu käyttöön kesällä 2021.

Mallinnuksen tavoitteita ja kehitystyötä on dokumentoitu ECCOLA-menetelmän avulla. ECCOLA-menetelmän tarkoituksena on ohjata tekoälymallinnuksen eri sidosryhmiä dokumentoimaan tekoälyn tavoitteita, järjestelmällisiä ja eettisiä näkökulmia. ECCOLA-dokumentaatiossa pohditaan esimerkiksi mallinnuksen yhteiskunnallisia vaikutuksia (kuvio 16).

Wellbeing
#17 Societal Effects

Notes

Motivation: The impacts of a system go beyond its user-base. A system may affect negatively even those who do not use it nor wish to use it.

What to Do: Ask yourself:

- i. Did you assess the broader societal impact of the AI system's use beyond the individual (end-)users? Consider stakeholders who might be indirectly affected by the system.
- ii. How will the systems affect society when in use?
- iii. What kind of systemic effects could the system have?

Practical Example: Surveillance technology utilizing facial recognition AI has long-reaching impacts. People may wish to avoid areas that utilize such surveillance, negatively affecting businesses in said area. People may become stressed at the mere thought of such surveillance. Some may even emigrate as a result.

CID 8202-20201007

Date: 12.5.2021

Sprint: Final


Project week: [Click or tap here to enter text.](#)

Requirements related to the card

Mallilla on myös yhteiskunnallista vaikutusta, koska sillä pyritään ehkäisemään opiskelijoiden keskeyttämistä. Tavoitteena on, että koulutuksen järjestäjä pystyy tehostamaan omaa toimintaansa ja tunnistamaan vaikeuksia kokevat opiskelijat aikaisemmassa vaiheessa.

Notes

Voiko mallilla olla jotain negatiivista vaikutusta opettajien käyttäytymiseen liittyen? Jos opiskelijalle ennakoidaan keskeyttämistä, voiko ennuste alkaa toteuttamaan itseään (opettaja ei enää ohjaa opiskelijaa)? Voidaanko tämä estää jollain tavalla.



KUVIO 16 Tekoälymallinnuksen yhteiskunnalliset vaikutukset dokumentoituna ECCOLA-kortilla.

Toteutettu koneoppimisen malli on luotu opetusaineiston perusteella. Opetusaineistona on käytetty opiskelijatietoja vuoden 2018 alusta kuluvaan päivämäärään asti. Muistiinpanojen perusteella järjestelmiin kirjattujen opiskelijatietojen muoto on muuttunut reformin aiheuttamien muutosten takia merkittävästi, joten vanhempaa aineistoa ei ole ollut järkevää pitää mallinnuksessa mukana. Toteutetun mallin aineistossa on hyödynnetty muun muassa seuraavia tekijöitä

- tutkintotyyppi,
- tutkinto,
- valintatapa,
- opiskeluaika,
- ikä,
- pohjakoulutus,
- suoritukset,
- poissaolot sekä
- työssäoppimisen jaksot.

Rakennettu mallinnus käsittelee vain opiskelijoita, joiden koulutuksen tavoitteena on koko tutkinnon suorittaminen. Muistiinpanojen perusteella tämän tarkoituksena on ollut selkeyttää mallia käsittelemään pelkästään pitkäkestoisia koulutuksia, jolloin esimerkiksi lyhyet koulutukset, kuten

ensiapukurssit ja tulityökortit, jäävät aineiston ulkopuolelle. Näiden koulutuksien osalta tekoälyn mallinnuksella ei ole nähty lisäarvoa, eikä ne näin ollen myöskään monimutkaista rakennettua mallinnusta.

Mallin kehityksessä on otettu aineistosta 80 % opetusaineistoksi ja loput 20 % testiaineistoksi, jonka perusteella mallin suorituskykyä voidaan arvioida. Testiaineisto on mallille tuntematonta, eli sen sisältämiä havaintoja ei ole käytetty opetusaineistona. Käytetty aineisto on painottunut siten, että noin 80 % opetusaineistosta kuvaa valmistuneita opiskelijoita.

Mallinnuksessa on huomioitu opetukseen käytettävän opetusaineiston päivittyminen. Järjestelmistä siirtyy tietovarastoon päivittäin lisää aineistoa opiskelijoihin liittyen, mutta toteutettua tekoälymallia ei uudelleen kouluteta päivittäin. Mallin päivitystä ohjataan erillisillä ajoilla ja tällä hetkellä se on toteutettu manuaalisesti. Muistiinpanoissa huomioidaan, että jos suoritusten, poissaolojen tai muiden vastaavien asioiden osalta toimintatavat muuttuvat, tulisi malli uudelleen kouluttaa. Myös, jos tutkimukset tai muut opiskelijan taustatiedot päivittyvät, malli vaatii uudelleen kouluttamisen päivitettyllä opetusaineistolla. Tällainen tilanne voi tapahtua esimerkiksi silloin, jos aineistoon tulee uusi tutkinto.

Mallinnuksen sekavuusmatriisi on esitetty alla olevassa taulukossa (taulukko 8). Taulukon luvut ovat osuuksia kokoaineistosta, eikä varsinaisia havaintojen lukumääriä. Taulukossa positiivinen ennuste tai todellinen toteuma tarkoittaa opiskelijan valmistumista ja negatiivinen arvo kuvaa opiskelijan keskeyttämistä.

TAULUKKO 8 Mallinnuksen sekavuusmatriisi.

Opiskelija	Negatiivinen (ennustettu)	Positiivinen (ennustettu)
Negatiivinen (todellinen)	14,8 %	6,7 %
Positiivinen (todellinen)	1,5 %	77,0 %

Mallinnuksen tunnuslukuja on esitetty alla olevassa taulukossa (taulukko 9). Tunnusluvut on kerätty viimeisimmäksi tehdyn mallinnuksen arvioinnista. Tunnuksissa positiivinen ennuste ja toteuma tarkoittaa opiskelijan valmistumista.

TAULUKKO 9 Mallin tunnusluvut.

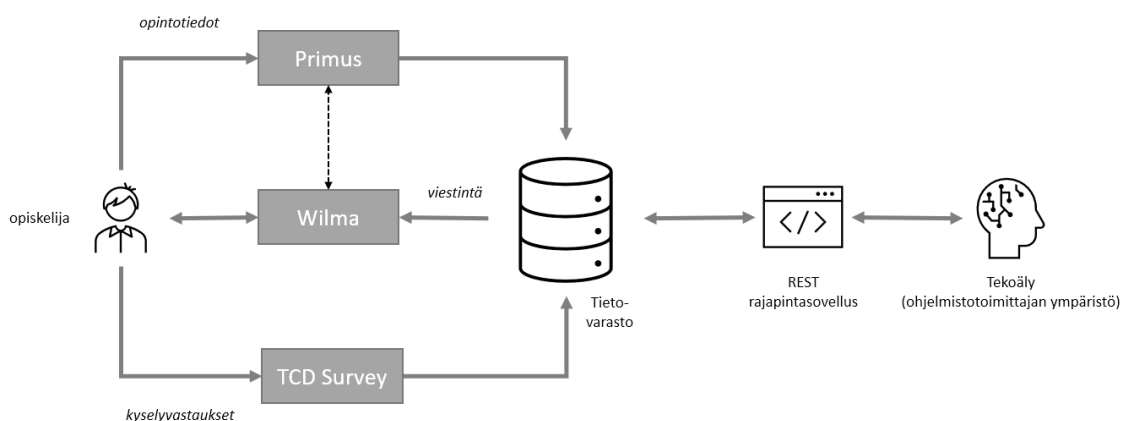
Tunnusluku	Arvo
Accuracy	0,919
Area under ROC-curve	0,964
True positive rate	0,981
True negative rate	0,692
False positive rate	0,308
False negative rate	0,019

Tietovarastoympäristöön on toteutettu ajastus, joka tekee päivittäin ennusteet opiskelijoille. Malli tuottaa ennustetun luokittelutiedon sekä lukuarvon arvoväliltä $[0, 1]$, joka kuvaa mallin tekemää tarkkuutta ennusteelle. Mitä lähempänä luku on arvopäitä, sitä vahvempi ennuste on.

Samiedulla on myös toisen kehittämishankkeen kautta käyttöönotettu tekoälyn toiminnallisuus, joka on syntynyt ”Älykäs ohjaus – tekoäly asiakaslähtöisessä opinto- ja uraohjauksessa” hankkeen tuloksena. Hanke on toteutettu vuosina 2020–2021 ja se on Kaakkois-Suomen ammattikorkeakoulun hallinnoima. Hankkeen tekniseen toteutukseen on osallistunut TCD Consulting and Research Oy sekä Kwork Innovations Oy, joka on toteuttanut hankkeen tekoälysovelluksen.

Hankkeen tarkoituksena on tarjota opiskelijalle henkilökohtaista apua opinto- ja uraohjaukseen. Järjestelmä viestittää opiskelijalle erilaisia työpaikka tai opinto-ohjaukseen liittyviä ehdotuksia. Ehdotusten taustalla on tekoälysovellus, joka luo ehdotukset opiskelijalle. Ehdotukset lähetetään opiskelijalle ajastetusti Wilma-viestinä.

Älykäs ohjaus -hankkeessa tarvittavat opiskelijatiedot haetaan Samiedun tietovarastosta. Aineiston tulkinta ja ehdotusten rakentaminen tapahtuu tekoälytoimittajan ympäristössä. Kokonaisuuteen liittyy myös palautteen kerääminen opiskelijoille lähetetyistä ehdotuksista (kuviot 17). Hankkeen materiaaleissa todetaan, että Samiedulla käytössä oleva tietovarasto on vauhdittanut hankkeen toteutumista, sillä lähdejärjestelmien tiedot voitiin suoraan viedä tekoälyn käsiteltäväksi rajapintasovelluksen avulla tietovarastosta.



KUVIO 17 Älykäs ohjaus -hankkeessa käytetty tietoarkkitehtuuri.

Projektiaineiston perusteella tekoälysovellus hakee avoimena olevat työpaikat ulkoisista palveluista. Työpaikkoja ehdotetaan opiskelijoille heidän opiskeluunsa liittyvien tietojen perusteella, kuten suoritettavan tutkinnon ja tutkintotutkimuksen. Ehdotusten hakeminen perustuu tekstin analysointiin (*NLP, Natural*

Language Processing). Tällöin esimerkiksi opiskelijan tutkintonimikkeelle ”auto-
asentaja” etsitään työpaikkailmoituksia, jotka vastaavat tätä hakusanaa.

Kappaleessa esitetyn aineiston perusteella voidaan muodostaa neljä ensisijaista empiiristä kontribuutiota (PEC3-6). Organisaatio on hyödyntänyt erilaisia kehityshankkeita ulkopuolisen teknisen asiantuntemuksen saamiseksi tekoälysovellusten kehittämiseen. Tekoälysovellukset käyttävät organisaation tietovarastoon jalostettuja tietoja lähdeaineistonaan.

PEC3: Organisaatio on osallistunut kehityshankkeisiin, joiden avulla se on saanut käyttöönsä tekoälyn sovelluksia.

PEC4: Tekoälyn ratkaisut ovat ulkopuolisten toimijoiden toteuttamia.

PEC5: Tekoälyn ratkaisut hakevat lähdeaineiston organisaation tietovarastosta.

PEC6: Päätöspuuna toteutetun koneoppimisen mallin tarkkuus on merkittävä. Mallin toimii organisaation omassa palvelinympäristössä.

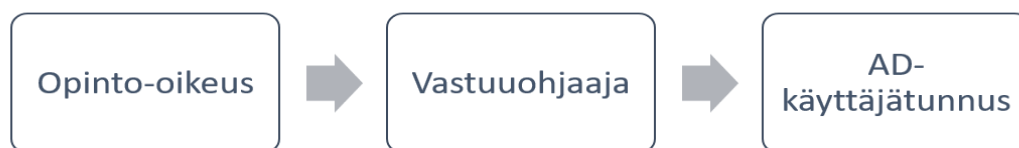
5.4 Tekoälyn hyödyntäminen

Opintojen etenemisen mallinnuksen tuottamaa aineistoa hyödynnetään Samie-
dussa vastuuhjaajan työpöydällä sekä PowerBI-raporteilla. Dokumentaatiossa on kuvattu ennusteen käyttäjäryhmiksi vastuuhjaajat, toimialajohtajat sekä koulutuksen järjestäjän johto.

Malli tuottaa tietoa oppilaitoksen omaan toimintaan (vastuuhjaajille, toimialajohtajille sekä johdolle) (ECCOLA kortti #13)

Vastuuhjaajan työpöytä on työkalu, jonka avulla järjestelmään kirjautuneelle käyttäjälle voidaan näyttää opiskelijatietoja hänen omalla vastuullaan olevista opiskelijoista. Työkalu on ollut käytössä vuodesta 2018 eteenpäin. Vastuuhjaajan työpöydällä esitetään opiskelijan ohjaukseen liittyviä tietoja, kuten opiskelijan opiskeluaika (aloituspäivä – arvioitu valmistumispäivämäärä), suoritettut osaamispisteet, HOKS-tapahtumat, erityisen tuen suunnitelmat, nykyinen opiskelijalaji, rahoitustiedot sekä työpaikalla järjestettävän koulutuksen sopimukset.

Vastuuhjaajan työpöydälle haettavien tietojen tietolähteenä toimii Samiedun tietovarasto, jonne on toteutettu tarvittavat tiedonsiirrot ja -haut. Vastuuhjaajan työpöytä vaatii kirjautumisen AD-käyttäjätunnuksilla. Kirjautuneen käyttäjätunnuksen perusteella haetaan siihen yhdistetyt opiskelijat tietovarastosta (kuvio 18).



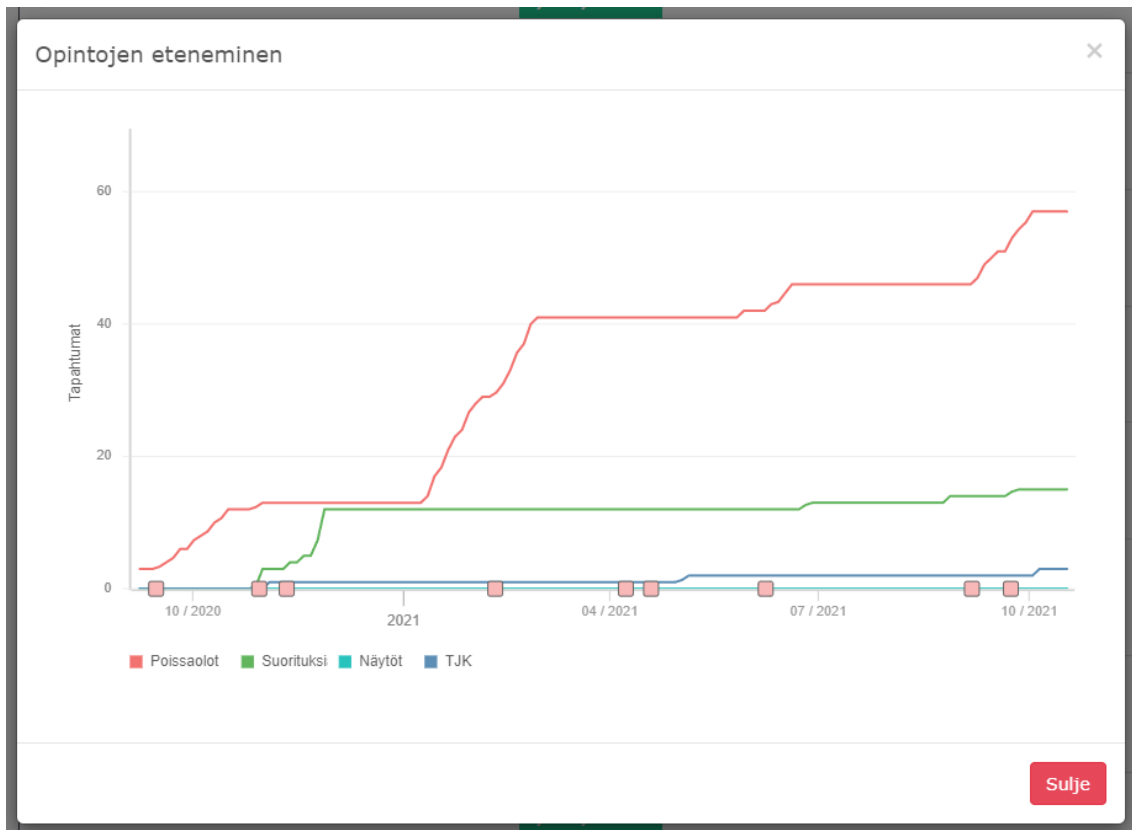
KUVIO 18 Käyttäjille voidaan esittää heidän omat opiskelijansa käyttäjätunnusten perusteella.

Vastuuohjaajan työpöydällä esitetään opiskelijakohtaisesti tekoälyn laskeuma ennuste. Ennusteen luokittelua ja pisteitystä ei esitetä sellaisenaan vastuuohjaajalle, vaan opiskelijan tila indikoidaan positiivisena tai negatiivisena. Indikaattori esittää keskeyttämishuonokkoa, jos ennuste on hyvin vahvasti keskeyttämisen puolella tai opiskelijan ennuste on muuttunut 30 päivän sisällä negatiiviseen suuntaan. Muissa tapauksissa indikaattori esittää positiivista opintojen etenemisen tilaa. Visuaalisesti tämä on esitetty punaisella tai vihreällä värillä korostamalla (kuvio 19).

Ryhmä	Nimi	Opiskelu aika	Osaamispisteet	HOKS	Erit.tuen tarve	Opiskelijalaji
Ryhmätunnus	Opiskelija 1 👍	05.08.2020 - 01.07.2022	Ammatilliset: 110,0 YTO: 35,0	Ensikertainen hyväksyntä OK! 27.08.2021	-	Tutkintoa suorittava
	Opiskelija 2 👍	05.08.2020 - 30.06.2023	Ammatilliset: 70,0 YTO: 22,0	Ensikertainen hyväksyntä OK! 10.08.2021	-	Tutkintoa suorittava
	Opiskelija 3 👍	05.08.2020 - 01.07.2022	Ammatilliset: 70,0 YTO: 35,0	Ensikertainen hyväksyntä OK! 27.08.2021	-	Tutkintoa suorittava
	Opiskelija 4 👎	07.09.2020 - 30.09.2023	Ammatilliset: 15,0 YTO: 13,0	Ensikertainen hyväksyntä OK! 23.09.2021	-	Tutkintoa suorittava
	Opiskelija 5 👍	07.09.2020 - 30.06.2023	Ammatilliset: 55,0 YTO: 0,0	Ensikertainen hyväksyntä OK! 27.08.2021	-	Tutkintoa suorittava

KUVIO 19 Esimerkki vastuuohjaajan työpöydästä, jossa esitetään tekoälyn ennuste vihreänä tai punaisena indikaattorina opiskelijan yksilöivän tiedon alapuolella (sarake 2).

Opiskelijakohtaisen ennusteen lisäksi vastuuhjaaja voi avata näkyville aikajanakuvaajan, jossa esitetään mallinnuksessa käytettyjä tietoja kuten opiskelijan poissaoloja ja suorituksia (kuvio 20).



KUVIO 20 Vastuuhjaajan työpöydältä avattavalla opiskelijakohtaisella aikajanakuvaajalla esitetään opiskelijan poissaolo ja suorituskertymiä.

Lisäksi samaa aineistoa hyödynnetään myös Samiedun toimialajohtajille ja muulle johdolle tarkoitetulla PowerBI-raportille. Raportin avulla voidaan tutkia opiskelijoiden etenemisen tilaa ja ennustetta muun muassa tutkintotyypeittäin ja tutkinnoittain. PowerBI on työkalu ja palvelu, jonka avulla voidaan visualisoida tietoa.

Ennustetietojen hyödyntämistä on pohdittu sen luotettavuuden perusteella dokumentaation muistiinpanoissa:

Jos järjestelmään automatisoi kyselyn mallin tietojen perusteella, voiko olla mahdollista, että väärä tulos aiheuttaakin negatiivisen kokemuksen opiskelijalle? (ECOLLA kortti #13)

Mallinnuksen auditointiin ja käyttöön liittyviä vaatimuksia on kuvattu dokumentaatioissa (kuvio 21).

Account-
ability

#18 Auditability

Motivation: Regulations affecting AI and data may necessitate audits of systems in the future. Similarly, if the system causes damage, an audit might be requested. It is good to have mechanisms in place beforehand.

What to Do: Ask yourself:

- i. Is the system auditable?
- ii. Can an audit be conducted independently?
- iii. Is the system available for inspection?
- iv. What mechanisms facilitate the system's auditability? How is traceability and logging of the system's processes and outcomes ensured?

Practical Example: In heavily regulated fields such as medicine, audits are typically required before a system can be utilized in the first place.



CID 9101-20200415

Notes

Date: 12.5.2021

Sprint: Final

Project week: [Click or tap here to enter text.](#)

Requirements related to the card

Järjestelmän tuottamat tiedot on voitava aikaleimata, jolloin voidaan todeta milloin ennusteita on tehty ja minkälaisilla arvoilla. Myös järjestelmän käyttämät lähtötiedot on voitava todentaa.

Järjestelmän aiheuttamat virheet (esim. epäonnistunut ajastus) pitäisi pystyä logittamaan ja lähettämään eteenpäin tiedoksi.

Järjestelmän tuottamien tietojen aikaleima tulisi olla myös näkyvillä loppukäyttäjillä, jolloin he tietävät million ko. ennusteet on tehty.

Notes

[Click or tap here to enter text.](#)

KUVIO 21 Mallinnuksen auditointiin liittyvät vaatimukset ja huomiot.

Aineiston perusteella ennustetiedon hyödyntäminen on vielä pilotointitasolla. Vastuuohjaajan työpöytä ei ole vielä käytössä koko henkilöstöllä, vaan vain osalla vastuuhjaajista. Muistioissa on pohdittu myös automaattisia hälytyksiä ennusteiden perusteella, mutta näitä ei ole otettu vielä käytäntöön. Muis-tion perusteella käyttöönotto vaatii vielä GDPR-vaatimusten selvittämisen au-tomaattisten viestien tai toimintojen toteuttamiseksi.

Älykäs ohjaus -hankkeen ehdotuksia lähetetään automatisoidusti opiskeli-joille heidän käyttämää Wilma-järjestelmään (kuvio 22). Ehdotuksia lähetetään vain opiskelijoille, jotka ovat antaneet siihen luvan Wilma-lomakkeella.

Wilma | Viestit | Työjärjestys | Oppimistehtävät / Tentit | ...

Oma etusivu > Viestit > Terveisiä tekoälyltä

Vastaustoiminnot | Toiminnot | Tulosta WWW-sivu

Terveisiä tekoälyltä

Lähtettäjä:	Botti
Vastaanottajat:	Meikäläinen Äly Seuraa lukeneita
Lähetetty:	22.4.2021 klo 13:00
Vastauksia:	Ei vielä yhtään

Tekoäly tuotti sinulle seuraavat ehdotukset opintoihisi liittyen:

Opintojen eteneminen
Oletko tyytyväinen opiskelutahtisi? Opiskelutahtisi on tällä hetkellä 45.5 osaamispistettä kuukaudessa:
- Yhteensä sinulla on 35 osaamispistettä.
- Se on 13 % tavoitenopeudestasi.
Ota yhteyttä ohjaajaasi, mikäli haluat apua opintojesi suunnittelussa tai edistämisessä.

Selite:
0-50 %, todella hidas
50-80 %, hidas
80-120 %, normaali
120-150 %, nopea
150 %, todella nopea.

Arvioi ehdotus: 🍌 / 🍌

Tähän viestiin ei voi vastata.

KUVIO 22 Tekoälyn muodostama ehdotus opiskelijan Wilmassa.

Viestin kuvauksessa on tekoälysovelluksen muodostama ehdotus, sekä lisäksi arviointilinkit, joiden avulla opiskelija voi antaa palautetta kyseisestä ehdotuksesta. Palautteessa annetaan arvio ehdotuksen hyödyllisyydestä sekä valinnainen avoin palaute. Palaute tuodaan tekoälyohjelmiston kehittäjän saataville, jolloin ehdotusten toimintalogiikkaa voidaan jatkokehittää.

Esitetyn aineiston perusteella voidaan todeta kaksi konkreettista tekoälyn sovellusta, joista voidaan muodostaa ensisijaiset empiiriset havainnot (PEC7 ja PEC8). Kolmantena merkittävänä havaintona voidaan todeta tekoälyn eettinen kehitystyö, sillä se on huomioitu molemmissa tekoälysovelluksissa (PEC9).

PEC7: Tekoälyn sovelluksen avulla voidaan ennustaa opiskelijoiden etenemistä sekä keskeyttämisriskiä.

PEC8: Tekoälyn avulla voidaan automatisoidusti lähettää opiskelijakohtaisesti mm. työpaikkaehdotuksia.

PEC9: Tekoälyn kehittämisessä on otettu huomioon eettiset näkökulmat. Esimerkiksi ECCOLA-menetelmän avulla dokumentoidaan tekoälysovelluksen kehitystyötä ja käyttöä.

5.5 Yhteenveto

Tutkimuksen tuloksena syntyi yhdeksän ensisijaista empiiristä kontribuutiota (PEC), jotka ovat koottuna alla olevassa taulukossa (taulukko 10). Tutkimuksen tuloksissa osoitettiin, kuinka kohdeorganisaatio on saanut käyttöönsä kaksi eri tekoälysovellusta, jotka ovat eri yritysten toimittamia. Oleellisena osana tekoälyn toteutuksia on organisaatioon toteutettu tietovarastoratkaisu, jonka tarkoituksena on koota useamman lähdejärjestelmän aineisto sekä tuoda se helpommin hyödynnettäväksi organisaation omassa toiminnassa ja kehittämisessä.

TAULUKKO 10 Aineiston pohjalta muodostetut empiiriset havainnot.

Tunniste	Empiirinen kontribuutio
PEC1	Organisaatiolla on eri käyttötarkoituksia varten useita operatiivisia ohjelmistoja, joihin muodostuu aineistoa.
PEC2	Tietovaraston avulla voidaan yhdistää useamman lähdejärjestelmän aineistot keskitettyyn tietokantaan. Integraatioiden avulla voidaan muokata lähdejärjestelmien tuottamaa aineistoa käsiteltävämpään muotoon, esimerkiksi raportointia varten.
PEC3	Organisaatio on osallistunut kehityshankkeisiin, joiden avulla se on saanut käyttöönsä tekoälyn sovelluksia.
PEC4	Tekoälyn ratkaisut ovat ulkopuolisten toimijoiden toteuttamia.
PEC5	Tekoälyn ratkaisut hakevat lähdeaineiston organisaation tietovarastosta.
PEC6	Päätöspuuna toteutetun koneoppimisen mallin tarkkuus on merkittävä. Mallin toimii organisaation omassa palvelinympäristössä.
PEC7	Tekoälyn sovelluksen avulla voidaan ennustaa opiskelijoiden etenemistä sekä keskeyttämisriskiä.
PEC8	Tekoälyn avulla voidaan löytää opiskelijakohtaisesti mm. työpaikkaehdotuksia. Ehdotukset voidaan välittää opiskelija automatisoidusti.
PEC9	Tekoälyn kehittämisessä on otettu huomioon eettiset näkökulmat. Esimerkiksi ECCOLA-menetelmän avulla dokumentoidaan tekoälysovelluksen kehitystyötä ja käyttöä.

6 Keskustelu

Tässä kappaleessa käydään läpi tutkimuksen tulokset, joita verrataan tekoälykyvykkyyden viitekehukseen sekä aiemmin esitettyyn kirjallisuuteen. Tulosten keskusteluosio jakaantuu kahteen kappaleeseen, käytännölliseen sekä tutkimukselliseen merkitykseen.

6.1 Käytännön merkitys

Tutkimus tuotti käytännöllistä tietoa siitä, kuinka tekoälyä voidaan hyödyntää ammatillisessa koulutuksessa koulutuksen järjestäjän näkökulmasta. Tutkimus osoitti tapaustutkimuksen avulla, kuinka opintohallintojärjestelmiin muodostuvaa aineistoa voidaan hyödyntää tekoälymallin toteutuksessa.

Kirjallisuusosiossa esitettiin aiempaa tutkimustietoa siitä, kuinka koulutuksen järjestäjä voi hyödyntää data-analytiikkaa tai tekoälyä omassa toiminnassaan. Näitä ratkaisuja ovat esimerkiksi

- Opiskelijoiden ohjeistusten kehittäminen (Ferguson, 2012; Koedinger, 2015)
- Opiskelijan ohjaaminen ja tutorointi (Ferguson, 2012; Koedinger, 2015)
- Opiskelijan säilyttäminen oppilaitoksessa (Kucak, 2018)
- Tulosten arviointi, testaaminen ja kokeet (Kucak, 2018)
- Tehokkuuden ja menestyksen ennustaminen (Ciolacu ym., 2017; Kucak, 2018; Kovac, 2010; Yess, 2009)
- Verkkoalustojen kehittäminen (Koedinger, 2015; Markellou ym., 2005; Romero & Ventura, 2007)

Tutkimuksessa osoitettiin, että kohdeorganisaatio oli omassa toiminnassaan toteuttanut lähdetietojärjestelmistä integraatiot tietovarastoon. Tietovarastoon siirretään aineistoa ajastetusti. Lähdejärjestelmien muodostamat siirtotiedostot eivät olleet suoraan hyödynnettävissä analytiikkatyökaluissa, vaan ne vaativat

myös toimenpiteitä, että tiedot saadaan ladattua tietokantaan hyödynnettäväksi. Tällaisia toimenpiteitä oli esimerkiksi tietotyyppien muunnokset sekä lähdejärjestelmän sisäisten listojen lataaminen tietokantaan. (PEC1 ja PEC2).

Kohdeorganisaatio oli hankkinut tekoälyn ratkaisua varten osaamisen ulkopuolelta (PEC3 ja PEC4). Tekoälyn mallinnuksen muodostaminen vaatii asiaan perehtymistä, sekä aineiston muodostamista oikeaan muotoon. Tekoälyn sovellusten lähdeaineisto on haettu organisaation tietovarastosta, jonka voidaan katsoa olevan keskeinen toiminto tekoälyn sovellusten toteuttamiseksi (PEC5).

Organisaation toiminta- ja laatujärjestelmään on muun muassa nostettu tietovaraston hyödyntäminen keskeiseen rooliin. Tulosten perusteella voidaan todeta, että kohdeorganisaatio on ottanut tiedolla johtamisen sekä tietovaraston ja sen ympärille kuuluvat työkalut osaksi omaa strategiaansa. Organisaatio on saanut myös tämän johdosta myös opetushallituksen laaturahaston vuonna 2017.

Tekoälysovelluksen avulla koulutuksen järjestäjä pyrkii hakemaan hyötyjä opiskelijan ohjaustoimintaan (PEC7). Tutkimuksen ajankohtana organisaatio oli ottanut pilotointikäyttöön tekoälysovelluksen, jonka avulla voidaan tukea vastuuhjaajan päätöksentekoa. Sovelluksen avulla pystytään tunnistamaan mahdollisessa keskeyttämisriskissä olevat opiskelijat ennakoivasti, ja tähän voidaan puuttua. Tutkimuksen tuloksena osoitetaan, että erilaiset näkymät voivat tarjota tämän informaation vastuuhjaajalle yksinkertaisessa muodossa, jonka lisäksi käyttäjällä on mahdollisuus tarkastella taustalla vaikuttavia tekijöitä vielä erikseen. Organisaation strategiassa myös todetaan, että reformin muutokset edellyttävät koulutuksen järjestäjältä rakenteiden ja prosessien rohkeaa uudistamista.

Lisäksi kohdeorganisaatiossa on otettu käyttöön opiskelijoille kohdennettu sovellus heidän opinto- ja uraohjaukseensa liittyen (PEC8). Sovelluksen avulla opiskelijat saavat automatisoidusti tekoälyn tekemiä ehdotuksia Wilma-viestinä liittyen heidän opintojensa etenemiseen sekä myös mahdollisesti sopivia työpaikkaehdotuksia. Tämän voidaan katsoa hyödyntävän opiskelijan henkilökohtaista opintopolkua.

6.2 Tutkimuksellinen merkitys

Tutkimuksen tuloksia voidaan verrata tutkimuksen teoreettiseen viitekehykseen. Mikalefin ja kollegoiden (2019) viitekehys kuvaa ulottuvuudet, joiden avulla organisaatio voi saavuttaa tekoälykyvykkyyden. Tutkimuksen empiiristen havaintojen perusteella voidaan todeta, että kohdeorganisaatio täyttää viitekehyksen esittämät viisi ulottuvuutta:

- data ja infrastruktuuri (PEC1 ja PEC2),
- osaaminen ja tekoälykulttuuri (PEC3) sekä
- teknologia (PEC5 ja PEC6).

Lisäksi organisaatio on hyödyntänyt tekoälyn ratkaisuja viitekehyksen osoittamalla tavalla päätöksen teon tukena (PEC7) sekä automaattisissa toiminnoissa (PEC8). Tutkimuksen tuloksena syntyneitä empiirisiä tuloksia verrataan kirjallisuudessa esitettyyn aineistoon alla olevassa taulukossa (taulukko 11).

TAULUKKO 11 Empiiristen havaintojen vertaaminen esitettyyn kirjallisuuteen.

Tunniste	Empiirinen kontribuutio	Suhde aiempaan kirjallisuuteen
PEC1	Organisaatiolla on eri käyttötarkoituksia varten useita operatiivisia ohjelmistoja, joihin muodostuu aineistoa.	Tukee kirjallisuutta (Kimball & Ross, 2011; Vassaliadis, 2002).
PEC2	Tietovaraston avulla voidaan yhdistää useamman lähdejärjestelmän aineistot yhtenäiseen tietokantaan. Integraatioiden avulla voidaan muokata lähdejärjestelmien tuottamaa aineistoa käsiteltävämpään muotoon, esimerkiksi raportointia varten.	Tukee kirjallisuutta (Kimball & Ross, 2011; Miloslavskaya & Tolstoy, 2016; Pyle 1999; Vassaliadis, 2002)
PEC3	Organisaatio on osallistunut kehityshankkeisiin, joiden avulla se on saanut käyttöönsä tekoälyn sovelluksia.	Tukee osittain kirjallisuutta. Opetus- ja kulttuuriministeriö esittää, että koulutuksen järjestäjien analytiikan työkalut voivat olla kolmannen osapuolen kehittämiä (Opetus- ja kulttuuriministeriö, 2021).
PEC4	Tekoälyn ratkaisut ovat ulkopuolisten toimijoiden toteuttamia.	Tukee kirjallisuutta. Mikalef ja kollegat (2019) esittävät, että organisaatio voi hankkia tekoälyyn liittyvät teknisen osaamisen ulkopuolelta.
PEC5	Tekoälyn ratkaisut hakevat lähdeaineiston organisaation tietovarastosta.	Tukee kirjallisuutta (Vaisman & Zimányi, 2014).
PEC6	Päätöspuuna toteutetun koneoppimisen mallin tarkkuus on merkittävä. Mallin toimii organisaation omassa palvelinympäristössä.	Tukee osittain kirjallisuutta (Alturki ym., 2020). Tekoälyn sovellus ei kuitenkaan vaadi merkittävää laskentatehoa tai pilvikapasiteettiä toisin kuin Mikalef ja kollegat (2019) esittävät.
PEC7	Tekoälyn sovelluksen avulla voidaan ennustaa opiskelijoiden etenemistä sekä keskeyttämisriskiä.	Tukee kirjallisuutta (Ciolacu ym., 2017; Kovac, 2010; Kucak, 2018; Yess, 2009)
PEC8	Tekoälyn avulla voidaan löytää opiskelijakohtaisesti mm. työpaikkaehdotuksia. Ehdotukset voidaan välittää opiskelija automatisoidusti.	Tukee kirjallisuutta (Araujo ym., 2020; Romero & Ventura, 2007).
PEC9	Tekoälyn kehittämisessä on otettu huomioon eettiset näkökulmat. Esimerkiksi ECCOLA-menetelmän avulla dokumentoidaan tekoälysovelluksen kehitystyötä ja käyttöä.	Tukee kirjallisuutta (Bostrom & Yudkowsky 2014; Vakkuri ym., 2021).

Tekoälyyn, koneoppimiseen ja yleisesti data-analytiikkaan liittyy olennaisesti lähdeaineisto ja sen valmistelu (Pyle, 1999). Aineiston käsittelyä varten

esiteltiin kirjallisuudessa erilaisia ratkaisuja, kuten tietovarastointi (*data warehouse*), tietotalvat (*data lake*) tai muut ”*big data*” ratkaisut (Miloslavskaya & Tolstoy, 2016). Tutkimuksen tulokset vahvistavat kirjallisuuden näkemyksen, että organisaatioiden tietoa, jota syntyy operatiivisista järjestelmistä, voidaan hyödyntää erilaisia analytiikan työkaluilla.

Tutkimuksen aiheesta ei ole tehty aiemmin vastaavia tutkimuksia Suomessa. Aiemmat tutkimukset, joita tämänkin tutkimuksen kirjallisuusosiossa käytiin läpi, liittyivät korkeakouluopintoihin tai ne oli toteutettu Suomen ulkopuolella. Tämän tutkimuksen tarkoituksena oli täyttää tämä tutkimusaukko. Vaikka tutkimus kohdistui toisen asteen ammatilliseen koulutukseen Suomessa, tuloksia on kuitenkin mahdollista hyödyntää myös muilla koulutusasteilla, huomioiden kuitenkin ammatilliseen koulutuksen liittyvät erityispiirteet.

Tutkimus tuotti uutta tietoa siitä, miten ammatillinen koulutuksenjärjestäjä voi hyödyntää tekoälyä omassa toiminnassaan. Tutkimuksen tuloksissa esitettiin ne asiat, joiden avulla se pystyi ottamaan käyttöönsä tekoälymallinnuksen. Lisäksi tutkimuksen tuloksissa esitettiin myös käyttötappauksia tämän informaation hyödyntämiseksi. Tutkimuksen tuloksena vahvistettiin Mikalefin ja kollegoiden (2019) viitekehyksen esittämät ulottuvuudet tekoälykyvykkyyden saavuttamiseksi. Lisäksi tutkimuksessa todettiin, että organisaatio hyödyntää tekoälyä omassa toiminnassaan päätöksen teon tukena sekä automatisoituna opiskelijaviesteinä.

Tulokset eivät osoita pääsääntöisesti ristiriitaisuuksia aiemmin esitettyyn olemassa olevaan tietoon, mutta kaikkia kirjallisuudesta esille nostettuja asioita ei tässä tutkimuksessa pystytty tai pyritty vahvistamaan. Kirjallisuudessa esitettiin, että esimerkiksi opiskelijan iällä ei ole merkitystä opintojen etenemiseen liittyen (Kovačić, 2010), mutta kohdeorganisaatiossa käytössä olevassa tekoälymallinnuksessa ikä oli yhtenä tekijänä.

7 Yhteenveto

Tämän tutkimuksen tavoitteena oli selvittää sitä, kuinka ammatillisen koulutuksen järjestäjä voi hyödyntää tekoälyä omassa toiminnassaan. Tutkimuksen lähtökohtana oli reformin tuomat muutokset koulutuksen järjestäjän toiminnassa, mikä on aiheuttanut koulutuksen järjestäjille muun muassa rahoituksen tiukentumista. Koska resurssit ovat vähentyneet sekä kuormittuneet, teknologisilta ratkaisuilta haetaan apua tähän ongelmaan. Yksi ratkaisu voi olla esimerkiksi tekoälyn avulla tehtävä automaatio tai sen tarjoama avustus päätöksenteossa.

Tutkimuksen varsinainen empiirinen osa toteutettiin 2021 kesällä tapaus-tutkimuksena Itä-Savon koulutuskuntayhtymään. Kohdeorganisaatio on ammatillisen koulutuksen järjestäjä, jonka vuosittainen opiskelijamäärä on noin 4 000 eri opiskelijaa. Tutkimuksessa hyödynnettiin Mikalefin ja kollegoiden (2019) teoreettista viitekehystä tekoälykyvykkyydestä ja kilpailukyvyistä, jonka esittämät ulottuvuudet tekoälykyvykkyyden saavuttamiseksi löydettiin myös tässä tutkimuksessa. Tutkimuksen aineisto kerättiin koulutuksen järjestäjän dokumentaatioista sekä tietojärjestelmien kautta tapahtuvalla havainnoinnilla. Organisaation uudistumisen strategiassa (2018-2021) on huomioitu reformin tuomat muutokset esimerkiksi opiskelijoiden henkilökohtaisempiin opintopolkuihin sekä rahoituksen muuttumiseen.

Koneoppimiseen ja tekoälyyn liittyviä aiempaa kirjallisuutta käytiin läpi tutkimuksen kirjallisuusosiossa. Tekoälyn ja koneoppimisen mallit pohjautuvat aiemmin havaittuun tietoon (Shalev-Shwartz & Ben-David, 2014). Mallien avulla voidaan löytää tiedosta esimerkiksi syy-seuraussuhteita (Pyle, 1999), poikkeamia (Hand, 2007), rakentaa ennusteita (Alturki ym., 2020) tai toteuttaa automaattisia toiminnallisuuksia (Araujo ym., 2020; Mikalef ym., 2019). Tekoälyn sovelluksilla katsotaan olevan yrityksen kilpailukykyä parantavia vaikutuksia (Davenport & Ronanki, 2018), sillä tekoälyn sovellukset voivat esimerkiksi toteuttaa manuaalisia tehtäviä ihmisiä nopeammin ja paremmin (Brynjolfsson ym., 2018). Tekoälyn sovellukset voivat myös toimia työn teon tukena, jolloin voidaan hyödyntää ihmisen ja koneen parhaita puolia (Jarrahi, 2018).

Data-analytiikkaan liittyy tiedon kerääminen, muuntaminen sekä jalostaminen, jota voidaan toteuttaa esimerkiksi tietovarastoinnilla (Kimball & Ross 2011; Vaisman & Zimányi 2014) tai tietoailla (Miloslavskaya & Tolstoy, 2016). Tietojen kerääminen voi tapahtua lähdejärjestelmistä, kuten oppilaitosten operatiivisista järjestelmistä tai muista tietolähteistä. Järjestelmiin voi kerääntyä alhaisen tason tapahtumatietoa, kuten opiskelijan tekemiä klikkauksia tai navigointeja sovelluksessa (Koedinger, 2015), järjestelmiin tallentuneita digitaalisia jälkiä (Larsson & White, 2014) tai se voi olla järjestelmiin kirjattuja tapahtumia, kuten opiskelijan suorituksia, poissaoloja tai osallistumistietoja (Alturki ym., 2020). Tämä tieto yhdistetään muuhun käytössä olevaan aineistoon, kuten opiskelijan demografisiin tietoihin.

Tekoälyyn ja koneoppimiseen käytettäviä algoritmeja ja päättelytapoja on useita, kuten esimerkiksi luokittelu (Shalev-Shwartz & Ben-David, 2014). Luokittelu voi tapahtua esimerkiksi päätöspuiden avulla, jotka tutkimuksen mukaan on yleinen koneoppimisen malleissa niiden käytettävyyden ja tehokkuuden takia (Alturki ym., 2020). Rakennettujen mallinnusten oikeellisuutta sekä tarkkuutta voidaan arvioida erilaisilla laskelmilla, kuten sekavuusmatriiseilla tai AUC-luvulla (Bradley, 1997).

Koneoppimisen ja tekoälyn sovellusten avulla voidaan rakentaa automatisoituja toimenpiteitä, tuottaa hyödyllistä informaatiota päätöksen teon tueksi, hyödyntää sitä markkinoinnissa tai innovaatioiden kehityksessä (Mikalef ym., 2019). Oppilaitosten opetushenkilöstölle voidaan esimerkiksi hakea opintovaikeuksia kokevat opiskelijat, jotka voivat tarvita lisäapua opintosuorituksissaan (Kučák ym., 2018). Kirjallisuudessa esitetään, että automaation avulla käytävissä olevat resurssit voidaan käyttää ja kohdentaa tällöin tehokkaammin (Brynjolfsson ym., 2018).

Tutkimuksessa empiirisen osion tuloksena todetaan, että kohdeorganisaatiolla on käytössä tietovarasto, jonka avulla on toteutettu muun muassa raportointinäkymiä useiden eri lähdejärjestelmien aineistoihin. Tietovaraston avulla on toteutettu kaksi erilaista tekoälyn sovellusta, joiden tarkoituksena on tukea opiskelijoiden opinto-ohjausta. Toinen tekoälyn sovelluksista on henkilökunnan käyttöön tarkoitettu opiskelijan opintojen etenemisen ennustaminen, joka esitetään vastuuhjaajien henkilökohtaisella työpöydällä. Etenemisen ennuste on toteutettu koneoppimisen mallilla, jossa malli luokittelee opiskelijan opintojen etenemisen positiiviseksi tai negatiiviseksi (keskeyttämishäikä).

Toinen tekoälyn sovellus on opiskelijoiden käyttöön tarkoitettu opinto- ja uraohjauksen työkalun. Työkalun avulla opiskelijalle voidaan lähettää henkilökohtaisia ehdotuksia opintojen etenemiseen tai mahdollisiin työ- ja harjoittelupaikkoihin liittyen. Työkalu hyödyntää luonnollisen tekstin analysoinnin menetelmiä, jonka avulla se pystyy luomaan opiskelijan tietoihin liittyviä ehdotuksia. Ehdotusten lähettäminen Wilma-viestinä on automatisoitu koulutuksen järjestäjän ympäristöön.

Tulosten sekä esitetyn kirjallisuuden perusteella voidaan todeta, että koulutuksen järjestäjillä on useita eri mahdollisuuksia tekoälyn käyttökohteiksi. Tekoälyn käyttöönottoon liittyy teknisiä näkökulmia, kuten tiedonhallinta, te-

koälyn mallin rakentaminen ja ylläpito. Organisaation osaamisen ja toimintakulttuurin näkökulmasta pitää taas pohtia, kuinka näitä työkaluja konkreettisesti hyödynnetään ja miten niillä saadaan tuotettua lisäarvoa. Tekoälyn sovellusten kehittämisessä ja käyttöönotossa tulisi pohtia eettisiä näkökulmia, kuten yksilön tietosuojaa. Organisaation toiminnan luonne voi rajoittaa tiettyjen toiminnallisuuden käyttöönottoa, kuten esimerkiksi automaattisia toimenpiteitä.

Tutkimuksessa esitettyyn ongelmaan, reformin myötä tiukentuneihin resursseihin, voidaan esittää yhtenä ratkaisuna tekoälyn sovelluksia ja niiden tuomia hyötyjä. Tutkimuksen tuloksissa todetaan, että tekoälyn sovelluksilla voidaan tuottaa liiketoimintaan lisäarvoa ennustetiedoilla sekä automaattisilla toimenpiteillä. Ennustetiedot toimivat organisaation päätöksen teon tukena. Automaattisilla toimenpiteillä taas voidaan saada hyötyjä, kun koneen tehtäväksi annetaan manuaalisten tehtävien suorittaminen. Näin ollen resurssit vapautuvat muuhun käyttöön.

7.1 Vastaukset tutkimuskysymyksiin

Tutkimuksen päätutkimuskysymys (1) voidaan jakaa kahteen osaan: mitä koulutuksen järjestäjältä vaaditaan tekoälyn ratkaisujen hyödyntämiseksi, sekä mitä konkreettisia tekoälyn ratkaisuja koulutuksen järjestäjä voi ottaa käyttöönsä. Tutkimuskysymystä tarkennettiin konkreettisemmaksi kahdella apukysymyksellä (2 ja 3).

1. *Miten koulutuksen järjestäjä voi hyödyntää tekoälyn ratkaisuja omassa toiminnassaan?*
2. *Kuinka tekoälymallinnus on pystytty rakentamaan?*
3. *Kuinka tekoälymallin tuottamaa tietoa hyödynnetään ja esitetään?*

Apukysymykseen 2. vastaus löytyy aiemmin esitetyn tutkimuksen teoreettisen viitekehyksen avulla, jota hyödynnettiin empiirisessä tutkimusosiossa. Tutkimuksen tuloksissa osoitettiin, että kohdeorganisaatiossa täytyi tekoälykyvykkyyteen liittyvät ulottuvuudet:

- Data
 - Organisaation operatiivisiin järjestelmiin muodostuu aineistoa niiden käytön myötä. Toteutetun tietovarastoratkaisun avulla operatiivisten järjestelmien tiedot on muokattu ja ladattu käyttökelpoiseen muotoon, jolloin aineisto on käsiteltävissä erilaisilla analytiikan työkaluilla.
- Teknologia
 - Mallinnukseen liittyvä teknologinen ratkaisu on ollut käytettävissä kohdeorganisaation ympäristössä omana tekoälysovelluksena. Lisäksi tekoälyteknologia on myös ulkoistettu ulkopuoliselle toimijalle.

- Infrastrukturi
 - Tietovarastoympäristö on tarjonnut rakenteet tekoälyn hyödyntämiseksi. Operatiivisiin järjestelmiin on olemassa tietoyhteyden rajapintojen toteuttamiseksi. Tietovaraston ja kohdeorganisaation infrastruktuurin avulla aineistoa on pystytty myös siirtämään ulkopuolisille toimijoille.
- Osaaminen
 - Kohdeorganisaation päätökset osallistua kehityshankkeisiin, joiden avulla organisaatio on saanut käyttöönsä asiantuntemuksen tekoälysovellusten toteuttamiseksi. Organisaatio on myös löytänyt tapoja hyödyntää tekoälyn kehittämää informaatiota.
- Tekoälykulttuuri
 - Organisaation strategiassa sekä toiminta- ja laatutoiminnassa on huomioitu tietovaraston tarjoama informaatio päätöksentekoa ja kehittämistä varten. Strategiassa myös mainitaan, että reformin muutokset edellyttävät rakenteiden ja prosessien uudistamista. Kehitystyö on ollut pitkäjänteistä. Organisaation osallistuminen tekoälyhankkeisiin niiden vieminen käyttöön asti osoittaa myönteistä asennetta tekoälyn sovelluksia kohtaan.

Apukysymyksen 3. vastaukseen voidaan todeta, että tekoälyn tuottamaa informaatiota hyödynnetään kohdeorganisaation vastuuhjaajien toiminnassa sekä organisaatiotasolla sitä voidaan hyödyntää päätöksen teossa erilaisten PowerBI-raporttien avulla. Tutkimuksen toteutuksen ajankohtana vastuuhjaajien käytössä oleva tekoälyn tuottama ennustetieto on vielä pilotointikäytössä. Lisäksi tutkimuksessa esitettiin opiskelijoille kohdennettu tekoälysovellus, jonka ehdotuksia opiskelijat pystyvät hyödyntämään henkilökohtaisesti omilla opintopoluissaan.

Näiden apukysymysten vastausten perusteella voidaan vastata tiivistetysti päätutkimuskysymykseen (1). Koulutuksen järjestäjä voi hyödyntää tekoälyn ratkaisuja toiminnassaan päätöksen teon tueksi, esimerkiksi tutkimuksessa esitetyn opiskelijan opintojen etenemisen ennustamisella. Esitetyn ennuste vaatii koulutuksen järjestäjältä panostusta infrastruktuuriin, teknologiaan sekä halukkuutta kokeilla erilaisia uusia teknologioita oman toiminnan kehittämiseksi. Lisäksi sovellukset voivat olla myös itsenäisiä sekä automaattisesti toimivia. Tutkimuksen kirjallisuusosiossa esitettiin myös muita vaihtoehtoisia toiminnallisuuksia tekoälyn sovelluksille, kuten automaattista opintosuoritusten arviointia tai kurssiehdotuksia perustuen opiskelijan suoritustietoihin.

7.2 Tutkimuksen heikkoudet

Tutkimuksen oikeellisuuden arviointi on laadullisessa tutkimuksessa hankalaa. Koska tutkimus toteutettiin tapaustutkimuksena, siinä tehdyt havainnot ja niistä pohjautuvat tutkimustulokset ovat tiettyyn tapahtuma-aikaan ja kontekstiin

riippuvaisia. Jos sama tutkimus toistetaan samassa organisaatiossa eri tutkijan toimesta, voi tutkimustulokset olla hieman erilaiset.

Remenyi, Williamsin, Moneyn ja Swartzin (1998) mukaan tapaustutkimuksia kohtaan voidaan esittää kritiikkiä, koska ne eivät välttämättä aina ole objektiivisia ja täsmällisiä. Objektiivisuudella tarkoitetaan, että todellisuus on tutkijasta riippumaton sekä jakamaton. Tapaustutkimuksessa tutkija on niin lähellä tutkittavaa kohdetta, että sitä voidaan pitää subjektiivisena. Toisaalta Farquharin (2012) mukaan tapaustutkimukseen liittyy kohteen tarkastelu, mutta sen tarkoitus on kuitenkin senhetkiseen kontekstiin kuuluvan ilmiön ymmärtäminen, eikä objektiivisuus ole tällöin tärkein tavoite. Tässä tutkimuksessa on pyritty täsmällisyyteen esittelemällä tutkimussuunnitelma, jossa on kuvattu tutkimusstrategia sekä tiedonkeruu ja analyysimenetelmät, joiden avulla tutkimus on toteutettu.

Voidaan myös todeta, että tämän tutkimuksen tekijä on tutustunut kohdeorganisaation ennen varsinaisen tutkimuksen aloitusta. Lisäksi tutkija on ollut osallisena toteuttamassa kohdeorganisaatiolle teknisiä ratkaisuja, jotka liittyvät tämän tutkimuksen aihealueeseen. Tämä läheisyys tutkijan ja tutkittavan kohteen välillä on voinut luoda liian optimistisen kuvan esitetyistä tuloksista. Darke ja kollegat (1998) esittävät, että tapaustutkimuksissa tutkijan omat taustat, näkemykset ja arvot voivat heijastua tuloksiin. Kuitenkin tutkimuksen tuloksissa tehdyt johtopäätökset perustuvat aineistoihin, havaintoihin ja dokumentaatioihin, jotka ovat myös muiden käytettävissä. Tällä tavalla tutkimuksessa on pyritty siihen, että havainnot ovat mahdollisimman neutraaleja ja objektiivisia.

Tapaustutkimuksiin voidaan esittää kritiikkiä myös sen tulosten huonon yleistettävyyden takia (Remenyi ym., 1998). Voidaan kuitenkin perustella, että tämän tutkimuksen tarkoituksaan ei ole luoda yleistettävää kuvaa siitä, miten koulutuksen järjestäjä voi hyödyntää tekoälyä, vaan tutkia tätä yksittäisenä otantana. Tämän tutkimus toimii siis esimerkkinä.

Tutkimuksessa käytettyä Mikalefin ja kollegoiden (2019) teoreettista viitekehystä tekoälykyvykkyydestä ja kilpailukyvyvystä ei hyödynnetty kokonaisuudessaan. Viitekehystä käytettiin havainnoimaan ne elementit, joita tarvitaan tekoälykyvykkyyden saavuttamiseksi, mutta viitekehysten esittämiä kaikkia hyötyjä sekä mahdollista kilpailukykyä ei tutkittu. Kohdeorganisaatiolla käytössä oleva tekoälyn mallinnus on käytössä vasta pilotointivaiheessa, joten kaikkia tapoja sen hyödyntämiseksi ei tässä vaiheessa ei ole vielä järkevää arvioida. Lisäksi nämä viitekehysten alueet eivät suoraan sopineet tämän tutkimuksen tutkimuskysymykseen.

Tutkimuksessa ei käsitelty yleisemmällä tasolla kohdeorganisaation toimintoja ja siellä tapahtuvaa mahdollista tekoälyn hyödyntämistä. Voidaan esimerkiksi todeta, että Samiedu käyttää markkinoinnissa digitaalisia palveluita, kuten sosiaalisen median markkinointiominaisuuksia mainosten kohdentamisessa tietyille kohderyhmälle. Tällainen mainosten kohdentaminen kuitenkin pohjautuu käytännössä tekoälyyn (Sterne, 2017). Tällaiset käyttötapaukset rajattiin tarkoituksella tutkimuksen ulkopuolelle tutkimussuunnitelmassa.

7.3 Jatkotutkimusaiheet

Tutkimuksen tulosten perusteella voidaan esittää useita eri aiheita jatkotutkimuksia varten. Aiemmassa kappaleessa todettiin, että Mikalefin ja kollegoiden (2019) viitekehystä ei hyödynnetty tässä tutkimuksessa kokonaisuudessaan. Jatkotutkimuksessa voitaisiin tutkia, mitä tekoälyn ratkaisuja koulutuksen järjestäjä voi käyttää viitekehysten muihin osa-alueisiin, kuin tämän tutkimuksen tuloksissa esitettyyn tuettuun päätöksen tekoon ja automaatioon. Muut viitekehysten esittämät osa-alueet ovat markkinointi sekä innovointi. Lisäksi tässä tutkimuksessa voitaisiin selvittää, onko tekoälyä hyödyntämällä koulutuksen järjestäjän mahdollista saavuttaa viitekehysten esittelemää kilpailuetua.

Tutkimuksen tuloksissa esitettiin tekoälyn mallinnus, joka pyrkii ennustamaan opiskelijoiden opintomenestystä. Ennusteen tarkkuutta on esitetty laskennallisilla mittareilla. Käyttönoton kannalta voitaisiin toteuttaa tutkimuksia siitä, miten tämä tekoälyn tuottama informaatio on vaikuttanut vastuuhjaajien toimintaan, sekä onko mahdollisia opiskelijoiden keskeyttämisä pystytty ennalkoimaan ja estämään tämän tiedon perusteella. Laskennalliset tunnusluvut ja mittarit perustuvat opetus- ja testausaineistoihin, mutta kuitenkin opintojen etenemisen ennusteen monimutkaisuuden takia voidaan esittää jatkotutkimusaihe, jonka tavoitteena olisi tutkia vastuuhjaajien tai opettajien subjektiivista näkemystä opiskelijoiden opintojen etenemiseen suhteessa tekoälyn tuottamaan ennusteeseen.

Tuloksissa myös todetaan, että nykyinen tekoälyn ratkaisu on tällä hetkellä vasta pilotointivaiheessa. Vastuuhjaajan työpöydälle lisättyä tekoälyn työkalun hyödyllisyyttä sekä käyttöönottoa voitaisiin tutkia käyttäjien näkökulmasta, eli nähdäänkö ennustetieto tarpeellisena ja onko siitä jotain konkreettista hyötyä opiskelijan ohjaukseen liittyen. Tällöin tutkimus voi selvittää myös sen, voiko työkalun avulla koulutuksen järjestäjä saavuttaa esimerkiksi resurssi-hyötyjä.

8 LÄHTEET

- Ali, S., Haider, Z., Munir, F., Khan, H., & Ahmed, A. (2013). Factors contributing to the students academic performance: A case study of Islamia University Sub-Campus. *American Journal of Educational Research*, 1(8), 283–289.
- Alturki, S., Hulpuş, I., & Stuckenschmidt, H. (2020). Predicting academic outcomes: A survey from 2007 till 2018. *Technology, Knowledge and Learning*, 1-33.
- Araujo, T., Helberger, N., Kruikemeier, S., & De Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & SOCIETY*, 35(3), 611-623.
- Bailis, P., Gan, E., Madden, S., Narayanan, D., Rong, K., & Suri, S. (2017, May). Macrobases: Prioritizing attention in fast data. In *Proceedings of the 2017 ACM International Conference on Management of Data* (pp. 541-556).
- Baker, R. S. J. d., & Yacef, K. (2009). *JEDM Journal of Educational Data Mining*, 1(1), 3-17.
- Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In *Learning analytics* (pp. 61-75). Springer, New York, NY.
- Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. *The Cambridge handbook of artificial intelligence*, 1, 316-334.
- Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30(7), 1145-1159.
- Brynjolfsson, E., Rock, D., & Syverson, C. (2019). 1. Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics (pp. 23-60). University of Chicago Press.

- Chen, M. S., Han, J., & Yu, P. S. (1996). Data mining: an overview from a database perspective. *IEEE Transactions on Knowledge and data Engineering*, 8(6), 866-883.
- Ciolacu, M., Tehrani, A. F., Beer, R., & Popp, H. (2017, October). Education 4.0—Fostering student's performance with machine learning methods. In 2017 IEEE 23rd international symposium for design and technology in electronic packaging (SIITME) (pp. 438-443). IEEE.
- Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*.
- Darke, P., Shanks, G., & Broadbent, M. (1998). Successfully completing case study research: combining rigour, relevance and pragmatism. *Information systems journal*, 8(4), 273-289.
- Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard business review*, 96(1), 108-116.
- Farquhar, J. D. (2012). *Case study research for business*. Sage.
- Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5-6), 304-317.
- Grønsund, T., & Aanestad, M. (2020). Augmenting the algorithm: Emerging human-in-the-loop work configurations. *The Journal of Strategic Information Systems*, 29(2), 101614.
- Hand, D. J. (2007). Principles of data mining. *Drug safety*, 30(7), 621-622.
- Heer, J. (2019). Agency plus automation: Designing artificial intelligence into interactive systems. *Proceedings of the National Academy of Sciences*, 116(6), 1844-1850.
- Hirsjärvi, S., Remes, P., & Sajavaara, P. *Tutki ja kirjoita*. 2015. 20. painos. Porvoo: Bookwell.
- Inmon, B. (2016). *Data Lake Architecture: Designing the Data Lake and avoiding the garbage dump*. Technics publications.
- Inmon, W. H. (1996). The data warehouse and data mining. *Communications of the ACM*, 39(11), 49-51.
- Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, 61(4), 577-586.

- Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1), 15-25.
- Kimball, R., & Ross, M. (2011). *The data warehouse toolkit: the complete guide to dimensional modeling*. John Wiley & Sons.
- Kovačić, Z. J. (2010). Early prediction of student success: Mining students enrolment data. In *Proceedings of informing science and IT education conference* (pp. 647-665).
- Kučak, D., Juričić, V., & Đambić, G. (2018). MACHINE LEARNING IN EDUCATION-A SURVEY OF CURRENT RESEARCH TRENDS. *Annals of DAAAM & Proceedings*, 29.
- Laine, M., Bamberg, J., & Jokinen, P. (2007). Tapaustutkimuksen käytäntö ja teoria. In *Tapaustutkimuksen taito* (pp. 9-38).
- Larose, D. T., & Larose, C. D. (2014). *Discovering knowledge in data: an introduction to data mining* (Vol. 4). John Wiley & Sons.
- Larusson, J. A., & White, B. (Eds.). (2014). *Learning analytics: From research to practice* (Vol. 13). Springer.
- LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436.
- Li, Y., Jiang, W., Yang, L., & Wu, T. (2018). On neural networks and learning systems for business computing. *Neurocomputing*, 275, 1150-1159.
- Lukkarinen, A., Koivukangas, P., & Seppälä, T. (2016). Relationship between class attendance and student performance. *Procedia: Social and Behavioral Sciences*, 228(July), 341-347.
- Ma, L.-C., & Wooster, A. R. (2009). Marital status and academic performance in college. *College Student Journal*, 13(2), 106-111.
- Markellou, P., Mousourouli, I., Spiros, S., & Tsakalidis, A. (2005). Using semantic web mining technologies for personalized e-learning experiences. *Proceedings of the web-based education*, 461-826.
- Mikalef, P., Fjørtoft, S. O., & Torvatn, H. Y. (2019, June). Developing an artificial intelligence capability: A theoretical framework for business value. In *International conference on business information systems* (pp. 409-416). Springer, Cham.
- Miloslavskaya, N., & Tolstoy, A. (2016). Big data, fast data and data lake concepts. *Procedia Computer Science*, 88, 300-305.

- Miloslavskaya, N., Senatorov, M., Tolstoy, A., & Zapechnikov, S. (2014, August). Information security maintenance issues for big security-related data. In 2014 International Conference on Future Internet of Things and Cloud (pp. 361-366). IEEE.
- Mohamadian, Z., Fallah, S., Safdarian, A., & Jalali, Z. (2015). An open access. *Indian Journal of Fundamental and Applied Life Sciences*, 5(S1), 1262–1270.
- Nithya, P., Umamaheswari, B., & Umadevi, A. (2016). A survey on educational data mining in field of education. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 5(1), 1–16.
- Noble, W. S. (2006). What is a support vector machine?. *Nature biotechnology*, 24(12), 1565-1567.
- Opetushallitus. (1.10.2021a). Ammatillisen koulutuksen laatupalkinnot. <https://www.oph.fi/fi/koulutus-ja-tutkinnot/ammattillisen-koulutuksen-laatupalkinnot>
- Opetushallitus. (1.10.2021b). Vuonna 2017 palkittujen palauteraportit. https://www.oph.fi/sites/default/files/documents/Vuonna%202017%20palkittujen%20palauteraportit_0.pdf
- Opetus- ja kulttuuriministeriö (2021). Oppimisanalytiikan viitekehys: Hyvät käytännöt oppimisanalytiikan käyttöön otossa ja hyödyntämisessä. Opetus- ja kulttuuriministeriön julkaisuja 2021:36.
- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, 30(3), 286-297.
- Provost, F., & Kohavi, R. (1998). Glossary of terms. *Journal of Machine Learning*, 30(2-3), 271-274.
- Remenyi, D., Williams, B., Money, A., & Swartz, E. (1998). *Doing research in business and management: an introduction to process and method*. Sage.
- Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1), 12-27.
- Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *ScienceDirect*, 33(33), 134–146.
- Samiedu. (1.9.2021). Uudistumisen strategia 2018-2021. https://samiedu.fi/wp-content/uploads/2020/10/SAMiedu_strategiaesite_2018-2021.pdf

- Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students*. Pearson education.
- Sengupta, S., Singh, A., Leopold, H. A., Gulati, T., & Lakshminarayanan, V. (2020). Ophthalmic diagnosis using deep learning with fundus images–A critical review. *Artificial Intelligence in Medicine*, 102, 101758.
- Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge university press.
- Shortliffe, E. H., & Sepúlveda, M. J. (2018). Clinical decision support in the era of artificial intelligence. *Jama*, 320(21), 2199-2200.
- Siau, K., & Yang, Y. (2017, May). Impact of artificial intelligence, robotics, and machine learning on sales and marketing. In *Twelve Annual Midwest Association for Information Systems Conference (MWAIS 2017)* (pp. 18-19).
- Soininen, J. (2021). *Tapaustutkimus-toimintamalli IT-johtamisen yhteydestä koulutusorganisaation tuloksellisuuteen koulutuksen reformissa*. JYU dissertations.
- Song, Y. Y., & Ying, L. U. (2015). Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry*, 27(2), 130.
- Sterne, J. (2017). *Artificial intelligence for marketing: practical applications*. John Wiley & Sons.
- Vaisman, A., & Zimányi, E. (2014). *Data warehouse systems. Data-Centric Systems and Applications*.
- Vakkuri, V., Kemell, K. K., Jantunen, M., Halme, E., & Abrahamsson, P. (2021). ECCOLA – A method for implementing ethically aligned AI systems. *Journal of Systems and Software*, 182, 111067.
- Vassiliadis, P., Simitsis, A., & Skiadopoulos, S. (2002, November). Conceptual modeling for ETL processes. In *Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP* (pp. 14-21).
- Watson, H. J., Goodhue, D. L., & Wixom, B. H. (2002). The benefits of data warehousing: why some organizations realize exceptional payoffs. *Information & Management*, 39(6), 491-502.
- Yassein, N. A., Gaffer, R., Helali, M., & Mohomad, S. B. (2017). Citation: Predicting student academic performance in KSA using data mining techniques. *Journal of Information Technology and Software Engineering*, 7(5), 213.

- Yess, J. P. (2009). Influence of marriage on the scholastic achievement of community college students: Humanities, social sciences and law. *American Journal of Educational Research*, 4(2), 103–118.
- Yin, R. K. (2009). *Case study research: Design and methods* (Vol. 5). sage.
- Yu-Wei, C. D. C. (2015). *Machine learning with R cookbook*. Packt Publishing Ltd.
- Zheng, N. N., Liu, Z. Y., Ren, P. J., Ma, Y. Q., Chen, S. T., Yu, S. Y., ... & Wang, F. Y. (2017). Hybrid-augmented intelligence: collaboration and cognition. *Frontiers of Information Technology & Electronic Engineering*, 18(2), 153-179.