



Kerttuli Ratilainen

Datan käsittelyn ja koneoppimisen ammattitaitovaatimukset ammatillisena työvoimakoulutuksena – pilotointi

Metropolia Ammattikorkeakoulu

Insinööri (AMK)

Tieto- ja viestintäteknikka

Insinöörityö

6.5.2024

Tiivistelmä

Tekijä:	Kerttuli Ratilainen
Otsikko:	Datan käsittelyn ja koneoppimisen ammattitaitovaatimukset ammatillisena työvoimakoulutuksena koulutuksena – pilotointi
Sivumäärä:	49 sivua + 1 liite
Aika:	6.5.2024
Tutkinto:	Insinööri (AMK)
Tutkinto-ohjelma:	Tieto- ja viestintätekniikka
Ammatillinen pääaine:	Ohjelmistotuotanto
Ohjaaja:	Lehtori Simo Silander
Ohjaaja:	ICT- ja mediaryhmän koulutuspäällikkö, Taitotalo, Elina Byckling

Tämän opinnäytetyön tavoitteena oli luoda työvoimakoulutus, joka soveltaa kansallisia datan käsittelyn ja koneoppimisen ammattitaitovaatimuksia ja joka vastaa työelämän tarpeisiin. Datan käsittely ja analysointi ovat tärkeitä taitoja digitaalisen murroksen aikana, ja Suomessa on datatalouden osaajien vaje. Koulutus on suunniteltu työttömille, työttömyysuhanalaisille ja vastavalmistuneille, joilla on jonkin verran kokemusta ohjelmoinnista, datan käsittelystä tai tilastomatematiikasta. Koulutus on tutkintoperustainen ja tähtää tieto- ja viestintätekniikan ammattitutkinnon osatutkinnon datan käsittely ja koneoppiminen suorittamiseen. Koulutus toteutettiin talvikaudella 2023-2024, se oli seitsemän kuukauden mittainen. Koulutuksen nimi on Edistyneen data-analytiikan osaaja. Koulutuksen toteutus eteni suunnitellusti.

Koulutuksen vaikuttavuutta arviotiin opiskelijoiden työllistymisen, harjoittelupaikan, sertifiointikokeiden, näyttötutkinnon ja opiskelijapalautteiden perusteella. Koulutuksen päätteeksi 42 % opiskelijoista työllistyi, löysi harjoittelupaikan tai aloitti rekrykoulutuksen, ja 12 opiskelijaa suoritti yhteensä 43 sertifiointikoetta. Opiskelijat kokivat oppineensa uusia taitoja koulutuksen aikana. Kehittävää palautetta tuli koulutuksen etenemistahdistista ja vaikeusasteesta. Koulutuksen aikana luotua opetussuunnitelmaa ja materiaaleja voidaan hyödyntää toiseen vastaavaan koulutukseen.

Avainsanat: datan käsittely, koneoppiminen, tekoäly, Python, työvoimakoulutus, ammatillinen koulutus

Tämän opinnäytetyön alkuperä on tarkastettu Turnitin Originality Check -ohjelmalla.

Abstract

Author: Kerttuli Ratilainen
Title: Adapting National Vocational Competence Requirements for Data Processing and Machine Learning to Labour Market Training – Pilot Project
Number of Pages: 49 pages + 1 appendices
Date: 12 May 2024

Degree: Bachelor of Engineering
Degree Programme: Information and Communication Technology
Professional Major: Software Engineering
Supervisors: Simo Silander, Senior Lecturer
Elina Byckling, Head of ICT and Media Team

The aim of this thesis is to show how to apply national vocational competence requirements for data processing and machine learning to labour market training skills, so that the training meets the needs of working life. Data processing and analysis are important skills during the digital transformation, and that there is a shortage of data experts in Finland. The training is designed for the unemployed, those threatened with unemployment and recent graduates who have some experience in programming, data processing or statistical mathematics. The degree programme is degree-based and aims at completing the data processing and machine learning as an optional unit of the vocational qualification in ICT. The training took place in winter 2023-2024 and lasted seven months. The name of the training is Advanced Data Analytics Expert. The training proceeded as planned.

The effectiveness of the training was evaluated based on student employment, certification exams, competence-based qualification, and student feedback. At the end of the training, 42% of the students found employment, internship or started recruitment training, and 12 students passed a total of 43 certification exams. The students felt that they had learned new skills during the training. Constructive feedback was given on the pace and difficulty of the training. The curriculum and materials created during the training can be used for another similar training.

Keywords: data processing, machine learning, artificial intelligence, Python, vocational education and training, labour market training

Sisällys

Lyhenteet

1	Johdanto	1
2	Insinööriyön tavoitteet ja menetelmät	2
2.1	Tavoitteet	2
2.2	Menetelmät	2
2.3	Aiemmat referenssit	3
3	Koulutuksen tarve (määrittely)	4
3.1	Taustaa	4
3.2	Kansalaisdatatieteilijöitä tarvitaan	6
3.3	Roolitus: datatieteilijä, datainsinööri vai data-analyytikko	8
3.4	Tutkinnon kuvaus (vaatimusmäärittely)	9
4	Koulutuksen suunnittelu, aikataulu ja opiskelijavalinta (suunnittelu)	11
4.1	Suunnittelu ja aikataulu	11
4.2	Opiskelijavalinta (asiakasvaatimukset)	12
5	Koulutuksen toteutus (toteutus)	15
5.1	Ensimmäinen osa: Johdanto	16
5.1.1	Johdanto-osuuden tavoitteet	17
5.1.2	Johdanto-osuuden toteutuksen kuvaus	18
5.1.3	Johdanto-osuuden opiskelijapalautteet	20
5.2	Toinen osa: Koneoppiminen, ohjattu ja ohjaamaton oppiminen	22
5.2.1	Ohjatun ja ohjaamattoman koneoppimisen tavoitteet koulutuksessa	23
5.2.2	Ohjatun ja ohjaamattoman koneoppimisen toteutus	24
5.2.3	Palautteet	26
5.3	Kolmas osa: Lisää työkaluja dataosaajalle: NLP, neuroverkot, tekoäly ja robotiikka, Git ja GitHub sekä tietovarastoyhteydet	27
5.3.1	Kolmannen dataosaajan työkalut osan tavoitteet	28
5.3.2	Kolmannen dataosaajan työkalut osan toteutus	28
5.3.3	Kolmannen dataosaajan työkalut osan palautteet	30
5.4	Neljäs osa: pilvipalveluiden perusteet, testaus ja automatisointi, MLOps, CI/CD	31

5.4.1	Pilvipalveluiden ja automatisoinnin tavoitteet	32
5.4.2	Pilvipalveluiden ja automatisoinnin toteutus	32
5.4.3	Pilvipalveluiden ja automatisoinnin palautteet	33
5.5	Työssäoppiminen, näyttö ja tutkinnon suorittaminen	34
5.5.1	Työssäoppimisjakson tavoitteet	34
5.5.2	Työssäoppimisjakson toteutus	34
5.5.3	Työssäoppimisjakson palautteet	36
6	Vaikuttavuuden arviointi (testaus)	36
6.1	Opiskelijoiden työllistyminen koulutuksen aikana jälkeen	36
6.2	Opiskelijoiden suorittamat sertifiointikokeet koulutuksen aikana	38
6.3	Opiskelijoiden antama palaute	39
6.4	Itseopiskelu	39
7	Johtopäätökset	40
8	Yhteenveto	43
	Lähteet	45
	Liitteet	
	Liite 1: Kuva: Citizen Data Science	

Lyhenteet ja termit

CI/CD:	<i>Continuous Integration</i> ja <i>Continuous Delivery/Deployment</i> . Ohjelmiston jatkuva integraatio ja jatkuva toimitus käynnissä olevaan tuotantoympäristöön.
Data-analytiikka:	Kuvailee ja analysoi tietoa ja tekee ennusteita. Edistyneessä data-analytiikassa on mukana automatisoituja toimintoja ja koneoppimista.
Datainsinööri:	Kerää, muokkaa ja tallentaa tietoa niin, että siitä saadaan toimiva muuhun käyttöön.
Datatiede:	Datan analysointia ja mallintamista tilastotieteellisin menetelmin sekä koneoppimisen keinoin.
ETL:	<i>Extract, Transform and Load</i> . Tiedon siirtämistä, muokkaamista ja tallentamista jatkokäyttöön. Osa datainsinöörin toimenkuvaa
Jupyter Notebook:	Avoimen lähdekoodin työkalu, joka sisältää sekä tekstiblokkeja että koodiblokkeja. Käyttö onnistuu verkkoselaimella. Soveltuu dokumentointiin ja kokeelliseen mallintamiseen. Tavallisimmin käytetään Python-ohjelmointikielellä, mutta se soveltuu myös muille ohjelmointikielille.
MLOps:	<i>Machine Learning Operations</i> . Kuvaa koneoppimismallien rakentamista ja kouluttamista, dataputken rakentamista ja versionhallintaa sekä tuotannossa olevan mallin monitorointia ja seuranta jatkuvan integraation ja jatkuvan toimituksen periaatteiden mukaan (ks. CI/CD).

Neuroverkot:	Neuroverkot ovat tekoälyn ala, jonka avulla pyritään käsittelemään ja analysoimaan suuria datamääriä. Niiden avulla analysoidaan erilaisia tietoja, kuten kuvaa tai kieltä.
NLP:	Natural Language Processing (NLP) eli luonnollisten kielten prosessointi, jonka avulla pyritään koneellisesti ymmärtämään ja käsittelemään luonnollista kieltä.
NumPy:	Numpy-kirjastoa käytetään Python-ohjelmoinnissa numeerisen datan käsittelyyn esimerkiksi tieteellisessä laskennassa, mutta myös valmistelemaan dataa visualisointeja varten.
Pandas:	Python-kirjasto datan muokkaukseen ja analysointiin.
Python:	Avoimen lähdekoodin monipuolinen ohjelmointikieli. Usein käytetty matemaattisten ongelmien ratkaisemiseksi datan käsittelyssä ja koneoppimisessa.
Sci-kit learn:	Koneoppimisen avoimen lähdekoodin kirjasto, jossa on kokoelma perinteisemmän koneoppimisen kirjastoja. Sivustolla on esiteltynä valmiita algoritmeja.
TensorFlow:	Koneoppimisen avoimen lähdekoodin kirjasto, joka on suunniteltu syväoppimisen tekoälysovellusten kehittämiseen.
TYPO-koulutus:	Työvoimakoulutus, jossa TE-palvelut tilaavat koulutuksen järjestettäväksi työttömille työnhakijoille.
Web Scraping:	Verkkoharavointi on verkosta löytyvien tietojen haravointia. Yksi käytetyimpiä python-kirjastoja on BeautifulSoup.

Visualisointi:

Datan visualisoinnissa pyritään yksinkertaistamaan monimutkaisia asioita ja tuomaan oleellisia asioita esiin. Tavallisimpia Python-kirjastoja ovat esimerkiksi Matplotlib ja Seaborn.

1 Johdanto

Tämä opinnäytetyö käsittelee datan käsittelyn ja koneoppimisen ammattitaitovaatimusten soveltamista ammatilliseen työvoimakoulutukseen. Kun koulutusta lähdettiin suunnittelemaan, ei vastaavaa toteutusta ollut julkaistu. Työn tuloksena syntyi opetussuunnitelma, joka määrittelee opetuksen tavoitteet, sisällöt, opetusmenetelmät ja arvioinnin perusteet osatutkintoon johtavaan ammatilliseen koulutukseen. Opetussuunnitelman tavoitteena on varmistaa, että opetus on tavoitteellista ja suunnitelmallista ja vastaa työelämän tarpeisiin. Opiskelijoille tulisi jäädä selkeä kuva, mitä heiltä odotetaan ja mitä he tulevat oppimaan. Ammatillisen osatutkinto liittyy tiettyyn ammattialaan tai erikoisalaan liittyvään osaamiseen.

Datan käsittelyn ja koneoppimisen opettaminen on tärkeää, koska digitalisaation muokkaa kaikkia aloja teollisuudesta hoivaan ja tietotyöhön. Käsiteltävän datan määrän kasvaminen on johtanut siihen, että tarvitaan asiantuntijoita ja tekijöitä, jotka osaavat hyödyntää olemassa olevaa dataa ja esimerkiksi yritysmaailmassa tehostaa olemassa olevia prosesseja ja tuoda datasta liiketoimintaetua yritykselle.

Tarkemmin motivaatio opinnäytetyön tekemiseen syntyi kouluttajan työstäni Taitotalossa, jossa tehtäväni oli suunnitella uusi työvoimakoulutus (TYPO-koulutus), joka vastaisi kasvaviin datan käsittelyyn ja analysointiin liittyviin tarpeisiin ja joka noudattaisi uudistettuja ammatillisia tutkintovaatimuksia [1]. Koulutuksen nimeksi tuli edistyneen data-analytiikan osaaja.

Tässä opinnäytetyössä kuvataan pilotoidun koulutuksen vaatimuksia ja toteutusta sekä toteutuksesta saatuja palautteita. Luvussa 1 on johdanto aiheeseen. Luvussa 2 käsitellään työn tavoitteita ja menetelmiä. Luvussa 3 määritellään työn tausta ja tarve sekä käsitellään teoreettista viitekehystä. Luvussa 4 avataan suunnitteluprosessia ja kerrotaan opiskelijavalinnoista. Luvussa 5 esitetään koulutuksen rakenne pääpiirteittäin eli kuvataan opetussuunnitelman tavoitteet, toteutus ja palautteiden kautta koulutuksen vaikutusta opiskelijoiden osaamiseen

motivaatioon ja työllistymisen. Luvussa 6 syvennyttään vaikuttavuuden arviointin. Luvussa 7 kerrotaan johtopäätökset, summataan työn tulokset ja pohditaan niiden merkitystä ja jatkokehitysmahdollisuuksia. Luvussa 8 on lyhyt yhteenveto.

2 Insinööriyön tavoitteet ja menetelmät

2.1 Tavoitteet

Työn tavoitteena on luoda uudelle koulutukselle opetussuunnitelma, joka perustuu kansallisiin ePerusteisiin, joka vastaa työelämän tarpeisiin ja joka tarjoaa työttömille työnhakijoille ajantasaista ja hyödyllistä tietoa. ePerusteissa määritellään valtakunnalliset tutkintoperusteet eri opetusasteille. Tieto- ja viestintätekniikan ammatillisen koulutuksen perusteita uudistettiin vuonna 2021 [1]. Koulutuksen tavoitteena on määrittää opetussuunnitelmalle sisältö, tavoitteet ja aikataulu eli toteutus, joka on käytännön läheinen ja joka tarjoaa työttömille työnhakijoille ajantasaista ja hyödyllistä tietoa. Koulutus tukee toteutuessaan työttömien opiskelijoiden työllistymistä.

Suunnitelmaa lähdettiin laatimaan pilottiprojektina tietäen toteutuksen olevan ensimmäinen laatuaan. Ensimmäisestä toteutuksesta saatujen kokemusten perusteella voidaan koulutusta kehittää jatkossa uudelleen toistettavaksi opetuskokonaisuudeksi.

2.2 Menetelmät

Pääaineeni on ohjelmistotuotanto. Ohjelmistotuotannon prosessiin kuuluu suunnittelu ja määrittely, toteutus, testaus sekä valmiin tuotteen ylläpito. Insinööriyöni tavoitteena on kuvata ohjelmistokehitysprojektin käsitteillä tieto- ja viestintätekniikan ammatillisen koulutuksen suunnittelu, toteutus ja vaikuttavuus.

Koulutuksen toteutus kuvataan tässä insinööriyössä ohjelmistoprojektin omainsena kehitysprojektina, jonka lopputuotteena syntyy ammatillinen tieto- ja

viestintätekniiikan osatutkintoon tähtäävä koulutus. Tässä insinööriyössä vastaavat otsikot ovat:

- koulutuksen tarve (määrittely)
- koulutuksen suunnittelu, aikataulu ja opiskelijavalinta (suunnittelu)
- koulutuksen toteutus (toteutus)
- koulutuksen vaikuttavuus (testaus).

2.3 Aiemmat referenssit

Koulutuksen suunnittelun ja toteutuksen taustalla ovat suorittamani kurssit Metropoliasissa. Taulukossa 1 nimetään syventävien ja valinnaisten opintojen moduulit, jotka ovat oleellisesti vaikuttaneet sekä tämän insinööriyön, että ammatillisen koulutuksen opintosuunnitelman luomiseen.

Taulukko 1. Opintosuunnitelman sisältöön oleellisesti vaikuttaneet opintomoduulit, niiden vastuuopettajat sekä niiden laajuus Metropoliasissa [2].

Syventävät ammattiopinnot	opettaja	op
Datan käsittely ja koneoppiminen TX00DZ36	Juha Kopu ja Vesa Ollikainen	5
Neuroverkot koneoppimissovelluksissa TX00EW91	Sakari Lukkarinen ja Juha Kopu	5
Todennäköisyyslaskenta ja tilastomatematiikka TX00CK75	Mikko Pere	5
Valinnaiset opintokokonaisuudet		
Introduction to R Programming TX00EV62	Erika Hagan Brown	3
Koneoppimisen ja neuroverkkojärjestelmien matematiikka ja menetelmät TX00DV61	Vesa Ollikainen ja Mikko Pere	5
Tekoälyn perusteet	Elements of AI	2
	yhteensä	25

3 Koulutuksen tarve (määrittely)

3.1 Taustaa

Sitran Megatrendit 2017 -raportissa mainitaan työelämän muutoksen vaikuttimina ”automatisaatio, robotisaatio, keinoäly ja digitaaliset alustat muokkaavat kaikkia aloja teollisuudesta hoivaan ja tietotyöhön” [3., kalvo 8/39]. Valtiovarainministeriön julkaisussa ’Pilkahduksia tulevaisuuteen - digitalisaation ja robotisaation mahdollisuudet’ luetellaan keksintöjä, jotka pikkuhiljaa mahdollistivat muutokset, esimerkiksi:

- henkilökohtaiset tietokoneet (PC:t) ja myöhemmin mobiililaitteet ja uudenlaiset päätelaitteet, kuten esimerkiksi VR-laitteet, sekä laitteiden kapasiteetin kasvaminen
- tekoälymallien kehittyminen
- tiedon käsittelyn automatisointi
- siirtyminen työpöytäsovelluksista web- ja mobiilisovelluksiin
- jaettu kapasiteetti esimerkiksi pilvipalvelujen muodossa. [4, s.1–19.]

Kehitys on mahdollistanut massatiedon keräämisen laitteista ja internetistä. Kaikesta toiminnasta jää merkintä vähintäänkin lokitiedostoon. Samalla 2010-luvulla jatkunut neuroverkkojen ja syväoppimisen kehitys edesauttoivat datan käsittelyn nopeutumista. Tietokoneiden muisti ja kapasiteetti käsitellä tietoa on kasvanut.

Käsiteltävän datan määrän kasvaminen on johtanut siihen, että tarvitaan asiantuntijoita, jotka osaavat hyödyntää olemassa olevaa dataa ja esimerkiksi yritysmaailmassa tehostaa prosesseja ja tuoda datasta liiketoimintaetua yritykselle. Datan merkitys ja hyödynnettävyys kasvaa koko ajan. Toisaalta nykypäivän digitaalisessa maailmassa olemme kaikki tietolähteitä. Käyttäjän toiminta tallentuu johonkin tietovarastoon, eikä aina ole selvää, kuinka paljon tietoa onkaan tallennettuna eri toimijoiden käyttöön ja miten niitä tietoja voi yhdistää.

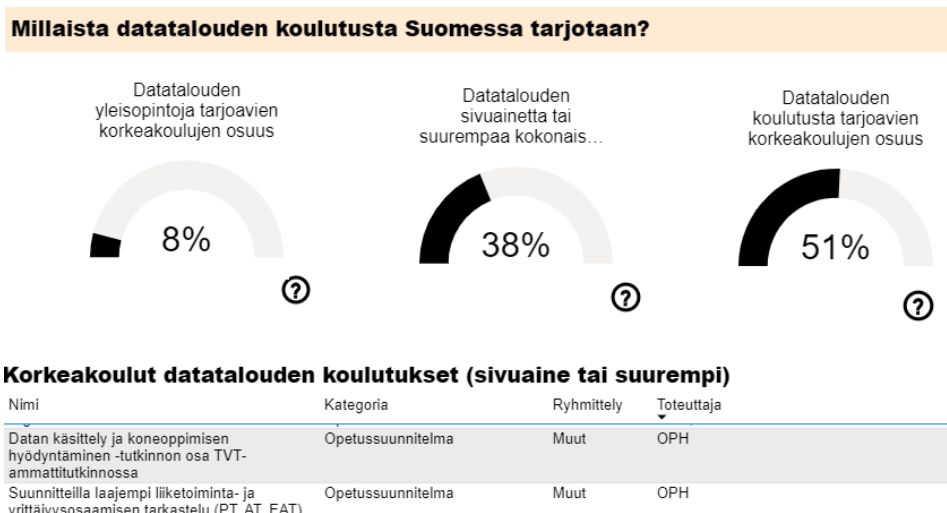
Sitra toteutti syyskuussa 2021 ’Datatalouden perusteet osaksi kansallistaitoja’ -kyselyn [5] peruskoulun ja toisen asteen opettajille tavoitteenaan tutkia opettajien asenteita ja osaamista liittyen datatalouteen. Kyselyssä painoi huoli kansalaisen

henkilökohtaisen datan turvasta. Kyselyn julkaisussa datatalous määritellään seuraavasti:

” Datatalous leikkaa läpi kaikki talouden alueet. Kaikilla toimialoilla syntyvästä datasta muodostuu raaka-ainetta, josta voidaan saada lisäarvoa. Yrityksen tai muun toimijan liiketoimintamalli määrittelee, millaisesta datasta ja millä keinoin lisäarvoa pyritään luomaan.” [5.]

Opetushallituksen (OPH) tietopaketti 'Datatalousosaamisen perusteita perusopetukseen ja toiselle asteelle' [6] todetaan, että ”data ei ole ”uusi öljy””. Tällä viitataan siihen, että vaikka datasta puhutaan välillä uusiutuvana luonnonvarana, näillä nykyisillä toimilla data ei tule häviämään samalla tavoin kuin luonnosta saatavat raaka-aineet. Tietopakettissa todetaankin, että dataa ”voidaan hyödyntää, jakaa ja jalostaa loputtomasti ilman, että se kuluu tai loppuu tai että sen arvo vähenee.”

Suomessa on Sitran tilannekatsauksen mukaan noin 15 000 dataosaajan vaje. Datan käsittelyä koulutetaan useissa korkeakouluissa ja muissa oppilaitoksissa ja koulutuksia kehitetään entisestään ”vastaamaan nykypäivän dataosaamisen tarpeisiin” [7]. Tilannekuvassa (kuva 1) on osaamisen välilehdellä päivitetty vuoden 2023 tilanne, eikä Taitotalon koulutusta näy vielä siinä [8].

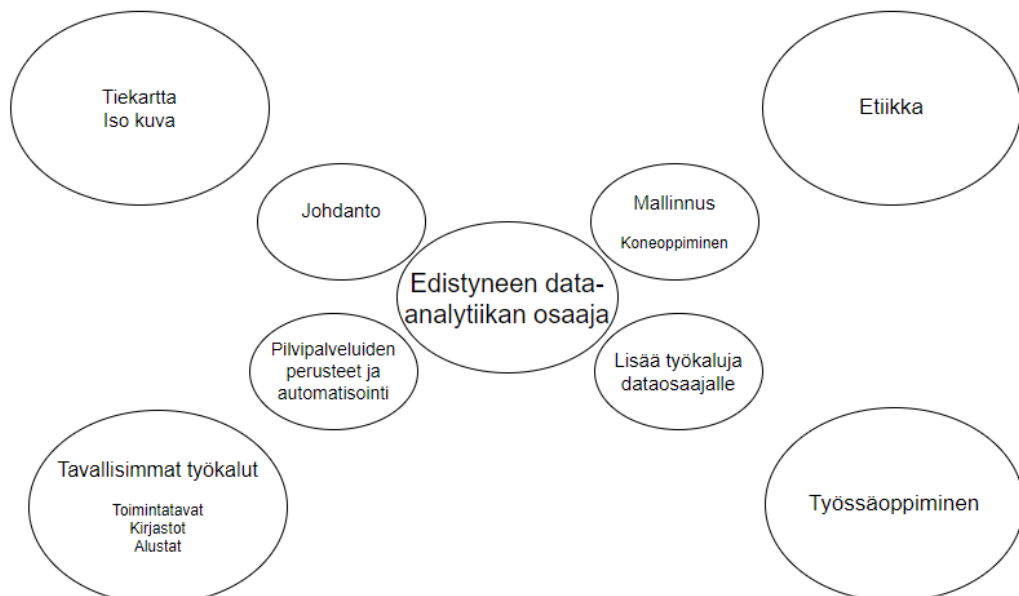


Kuva 1. Datatalouden tilannekuva – Power BI -raportti, osaaminen-välilehti [8].

3.2 Kansalaisdatatieteilijöitä tarvitaan

Joulukuussa 2022 Helsinki *Data Science Meetup* järjesti tapahtuman Datatieteen parhaat käytännöt (*Data Science Best Practices*). Tapahtuman tallenne on kuunneltavissa YouTubessa [9]. Tallenteen keskivaiheilla Osuuspankin vastaava datatieteilijä Juha Vesanto esittelee määritelmän kansalaisdatatieteestä (*citizen data science*) (liite numero 1). Kansalaisdatatiedettä harjoittaa henkilö, joka toimii datan parissa, mutta ei ole koulutukseltaan tai ammatiltaan datatieteilijä. Riittää kun kansalaisdatatieteilijällä on pääsy dataan, hän ymmärtää datatieteen perusteet ja hän voi luoda liiketaloudellista tai muuta arvoa datasta. Kansalaisdatatieteilijä voi olla kuka tahansa datan kanssa töitä tekevä, esimerkiksi ohjelmoija, taloustieteilijä, datainsinööri, matemaatikko jne.

Vesannon diaesitys inspiroi edistyneen data-analytiikan osaajan koulutuksen rakenteen ja sisällön määrittelyssä. Kuvan 2 kaavio mukailee Vesannon esittämiä datatieteen hyviä käytäntöjä. Niistä on valittu tämän koulutuksen kannalta olennaisia asioita. Kuvassa 2 on hahmoteltu taitoja ja työkaluja, joita tarvitaan nykyään datan käsittelyssä.



Kuva 2. Edistyneen data-analytiikan koulutuksen rakenne ja sisältö otsikoitain ja mukaillen lähteestä [9].

Kuvassa 2 isommat ympyrät ulkokehällä ovat koulutuksen läpileikkaavia teemoja ja sivuavat kaikkia koulutuksen aikana käytyjä aiheita. Vasemmalla ylhäällä tiekartta kuvastaa tässä tapauksessa sitä, että noin seitsemän kuukauden koulutuksessa ei päästä missään asiassa kovin syvälle, mutta tavoitteena on, että opiskelijoille piirtyy yleiskuva asian tiimoilta tai pohja, jonka päälle rakentaa syventävää osaamista oman kiinnostuksen mukaan. Toisin sanoen opiskelijalle tarjotaan yleiskuva siitä, missä tällä hetkellä mennään datan käsittelyn ja koneoppimisen saralla.

Vasemmalla alhaalla (kuva 2) tavallisimmat työkalut tarkoittavat koodieditoria, kirjastoa, visualisointityökaluja, pilvi-infrastruktuuria. Jokaisen teeman kohdalla karotitetaan siihen sopivia työkaluja ja menetelmiä. Työkaluihin kuuluvat myös tietyt matemaattiset mallit, koska jos haluaa ymmärtää eri sovellusten koneoppimisen malleja, on hyvä osata tilastotiedettä.

Kuvassa 2 oikealla ylhäällä on etiikka. Kaikkeen datan käsittelyyn liittyy eettisiä kysymyksiä, lainsäädäntöä ja määräyksiä. Ajan kuluessa kerätyssä datassa voi olla esimerkiksi historiallista vääristymää, jos näin on, miten se tunnistetaan. Kuka määrää, kenen päälle itseohjautuva robottiauto voi ajaa kolaritilanteessa. Koulutuksessa näihin kysymyksiin tutustutaan pinnallisesti, koska pääpaino on enemmän työkaluissa. Toisaalta jokaisen kansalaisen tulisi tietää, mitä oikeuksia hänellä on oman tietonsa suojaamiseen. Yleisten hyvien toimintatapojen ja säännösten tunteminen tulisi luoda raamit kaikelle, mihin dataa käytetään.

Oikealla alhaalla (kuva 2) on työssäoppiminen, koska koulutus tähtää ammatillisen osaamisen kehittymiseen ja työllistymiseen. Toisen asteen ammatillisessa opiskelussa pyritään, että oppiminen tapahtuu työn kautta ja tehden, enemmän kuin opettelemalla teoreettista viitekehystä. Tavoitteena on pitää teoreettinen viitekehys niin matalana kuin mahdollista ja oppia enemmän käytännönharjoitusten kautta. Tämä on tasapainoilua monimutkaisen aiheen yksinkertaistamisena niin, ettei mitään oleellista jätettä pois.

Neljä pienempää ympyrää kuvaavat koulutuksen toteutusta jaettuna neljään eri alaotsikkoon (kuva 2). Huolimatta siitä, että tässä jaottelussa ympyröitä on viisi, jaettiin koulutus ajallisesti neljään lähes yhtä suureen osaan:

1. johdanto ja visualisointi
2. mallinnus: koneoppimisen mallit: ohjattu oppiminen ja ohjaamaton oppiminen
3. lisää työkaluja dataosaajalle: NLP, neuroverkot, tekoäly, robotiikka, tietovarastot ja versionhallinta
4. pilvipalvelut ja automatisointi: pilvipalveluiden perusteet, testaus ja automatisointi, MLOps, CI/CD.

Luvussa viisi Koulutuksen toteutus käydään tarkemmin kyseisten osien sisältö.

3.3 Roolitus: datatieteilijä, datainsinööri vai data-analyytikko

Koulutuksen suunnittelun teoreettisena lähtökohtana oli ajatus kansalaisdatatieteilijästä, jolla on hyvä yleiskäsitys alasta, mutta ei vielä mitään erityistä syvempää osaamisalaa, paitsi oma ammattiosaamisensa joltain lähellä olevalta alalta. Aika nopeasti oppilaspalautteissa ilmeni kuitenkin halu ymmärtää konkreettisia rooleja ja tehtäviä. Joku opiskelijoista kommentoi, että koulutuksen nimi johtaa harhaan ja siksi haluan nostaa jo määrittelyvaiheeseen kolmen roolin erot ja mihin niistä koulutuksen aikana pyritään vastaamaan.

Datatieteilijän rooliin kuuluu esimerkiksi algoritmien ja tietomallien luominen sekä koneoppimisen soveltaminen. Lähimpänä tätä roolia olimme koulutuksen kakkos- ja kolmososissa käsitellessämme eri koneoppimisen menetelmiä.

Datainsinööri muokkaa dataympäristöstä toimivan. Datainsinööri kerää ja muokkaa tietoa analysoitavaan muotoon ja integroi analysoidut tiedot järjestelmän käyttöön. Työnkuvaa kuvastaa termi ETL: *Extract, Transform, Load*. Koulutuksessa Johdanto-osuuden tiedon muokkaaminen sekä kolmannen osan tietovarastoyhteydet lähestyivät eniten tätä osuutta.

Data-analyttikko kerää ja analysoi tietoa. Tiedon visualisointi ja raportointi on korostetumpi verrattuna kahteen edelliseen tehtävään. Data-analyttikon tehtävänä on kuvailla tietoa ja luoda esimerkiksi ennusteita, mutta työkaluina on useammin Excel tai Power BI kuin ohjelmointikielet. Koulutuksessa tätä osuutta käytiin eniten ensimmäisessä johdanto-osuudessa, missä opeteltiin visualisointeja Pythonin kirjastoja ja Power BI:tä käyttäen.

Huolimatta ennakkotiedoista osalle osallistujista tuli yllätyksenä, että data-analytiikkaa opiskeltiin vähiten koulutuksen aikana. On selvää, että vaikka kaikkia kyseisiä rooleja sivutaan koulutuksen aikana, ei missään niissä tulla ammattilaiseksi muutamassa kuukaudessa. Parhaimmillaankin annetussa ajassa voi saada perustan, jonka päälle voi alkaa lisäämään osaamista toivomaltaan osa-alueelta. Osin nämä roolit menevät myös työelämässä päällekkäin.

Kolmen edellisen roolin tai ammatin lisäksi on nousemassa neljäs ammattinimike nimittäin koneoppimisinsinööri (ML-insinööri), jonka tärkeimpiä taitoja on ohjelmointitaidot, koneoppimisen algoritmien ymmärtäminen, SQL ja operoiminen pilvipalvelussa (Cloud Computing) [10]. Tämä on seikka, johon kiinnitetään huomiota seuraavassa toteutuksessa.

3.4 Tutkinnon kuvaus (vaatimusmäärittely)

Taitotalossa tutkintoperustaiset koulutukset sidotaan valtakunnallisiin tutkintoperusteisiin, jotka ovat kuvattuna ePerusteissa [9]. Kaikki TYPO-koulutukset ovat tällä hetkellä tutkintoperustaisia, vaikka usein niissä on tarkoitus suorittaa vain yksi osatutkinto. Myös edistyneen data-analytiikan osaaja -koulutuksen toteutus on sidottu olemassa oleviin tutkintoperusteisiin.

Tieto- ja viestintätekniikan ammatilliset tutkinnot jaetaan perustutkintoon, ammattitutkintoon ja erikoisammattitutkintoon. Tutkintojen valtakunnallisia ammattitaitovaatimuksia uudistettiin vuonna 2021 [1], jolloin Tieto- ja viestintätekniikan ammattitutkinnon (entinen datanomi (AT)) valinnaisiin tutkinnon osiin lisättiin kymmenen uutta osaa. Yksi uusista tutkinnonosista on osa 14: Datan käsittely ja

koneoppiminen, jonka laajuus on 40 osaamispistettä (osp) [1]. Koko ammattitutkinnon laajuus on 150 osp.

Osatutkinnon hyväksytyyn suoritukseen kuuluu kaksi ammattitaitovaatimusta:

”Opiskelija osaa

- muodostaa datalla ratkaistavan ongelman ja käsitellä tietoa
- hyödyntää koneoppimista.”

Lisäksi OPH:n määrittelemissä ePerusteissa [9] listataan hyväksytyyn suoritukseen kriteerit:

” 1) Opiskelija muodostaa datalla ratkaistavan ongelman ja käsittelee tietoa:

- kommunikoi asiakkaan kanssa ratkaistavasta ongelmasta tai kehitystavoitteista
- muodostaa datalla ratkaistavan ongelman
- kerää ja yhdistää tietojärjestelmien dataa
- tarkistaa ja korjaa tietoa esimerkiksi vioittuneen tiedon, puuttuvien arvojen, yhdistämis- tai muiden virheiden varalta.

2) Opiskelija hyödyntää koneoppimista.

- huomioi koneälyn käytön eettiset seuraukset
- käyttää ja optimoi koneoppimisen työkaluja ja algoritmeja
- seuraa koneoppimisen tuloksien tarkkuutta
- visualisoi dataa
- kokoaa tulokset ja kommunikoi asiakkaan kanssa jatkotoimenpiteistä.”

Opiskelijalla on oikeus saada valmistavaa koulutusta edellä mainituista aiheista. Ammatillinen tutkinto suoritetaan näyttötutkintona, joten opiskelijan tulee lopussa arvioitavassa näytössä osoittaa osaamisensa hyväksytyjen suorituskriteereiden mukaisesti.

4 Koulutuksen suunnittelu, aikataulu ja opiskelijavalinta (suunnittelu)

4.1 Suunnittelu ja aikataulu

Yksittäisen seitsemän kuukauden koulutuksen toteuttamiseen tulee varata aikaa puolitoista vuotta. Taulukko 2 näyttää suunnittelun ja toteutuksen eri vaiheet kuukausittain:

Taulukko 2. Suunnittelun ja toteutuksen eri vaiheet.

vuosi	kuukausi	tehtävä
2022	marras	ideointia
	joulu	alustava suunnitelma valmiina
2023	tammi	ehdotus TE-palveluihin
	tammi-maalis	suunnittelua muun työn ohessa
	huhti	Te-palveluiden hyväksyntä
	huhti-heinä	haku auki kyselyihin vastaaminen suunnittelua
	heinä	haku päättyy hakijoiden esikarsinta
	elo	opiskelijavalinnat ja haastattelut
	syys-joulu	koulutus alkaa valmistava koulutuksen toteutus yksittäisten koulutuspäivien suunnittelu
2024	tammi	valmistava koulutus jatkuu
	helmi-maalis	työssäoppiminen
	huhti	arviointit ja koulutuksen päätös

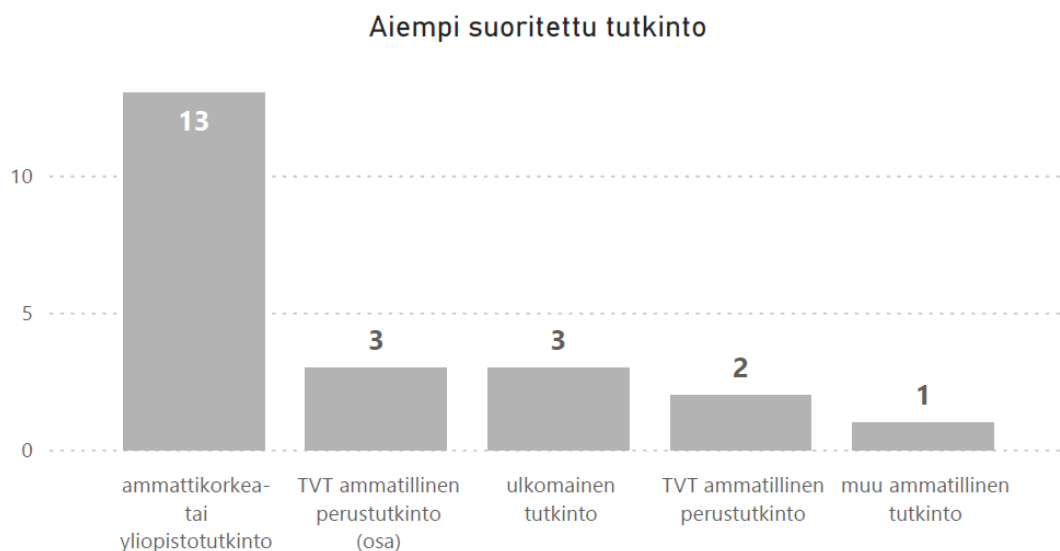
4.2 Opiskelijavalinta (asiakasvaatimukset)

Koulutukseen osallistuvia opiskelijoita voidaan pitää koulutustuotteen (sovelluksen) asiakkaina tai käyttäjinä. Yksi koulutuksen vaikuttavuuteen vaikuttavia tekijöitä on se, miten opiskelijat valitaan koulutukseen. Koulutuksen suunnittelun haasteena on, että ammatillisen koulutuksen tavoitteena on valmentaa opiskelijoita työelämään tarjoamalla käytännön taitoja, tietoja ja osaamista. Kun käsitellään esimerkiksi data-analytiikkaa tai koneoppimisen malleja, tarvitaan analyyttisiä taitoja, isojen kokonaisuuksien hahmottamista ja abstraktien käsitteiden hallintaa.

Koulutus on suunniteltu henkilöille, joilla on jonkin verran olemassa olevaa kokemusta ohjelmoinnista, ohjelmallisesta ajattelusta, datan analysoinnista, käsitteilystä ja/tai tilastomatematiikasta.

Työvoimakoulutuksen kohderyhmää ovat työttömät, työttömyysuhanalaiset, henkilöt, joiden työsuhde on päättymässä ja vastavalmistuneet. Hakijalla tulee olla soveltuvaa taustaa ja kokemusta joko työkokemuksen, koulutuksen tai harrastuksen myötä

Koulutukseen haki 176 työtöntä työnhakijaa TE-palveluiden kautta. Haastateluun valittiin 44 hakijaa, joista 22 valittiin koulutukseen. Kuvassa 3 havainnollistetaan osallistujien aiempaa koulutustaustaa. Ennen haastattelua hakijat täyttivät lomakkeen, jossa kysyttiin heidän taustatietojaan, motivaatiotaan sekä tarkistettiin omien laitteiden tekniset vaatimukset, esimerkiksi etäkokouksiin. Haastattelussa kerrottiin lyhyesti tulevasta koulutuksesta ja kartoitettiin opiskelijan motivaatiota ja taustaa.



Kuva 3. Osallistujien aiempi ylin suoritettu tutkinto. TVT-tarkoittaa tieto- ja viestintäteknikkaa.

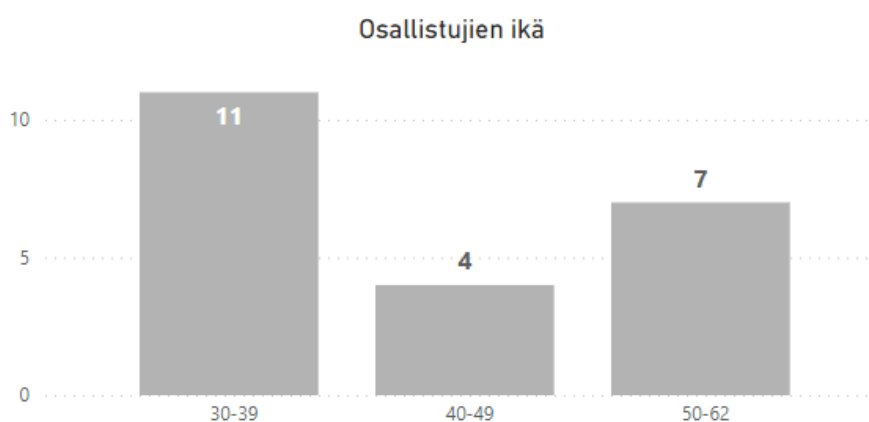
Kuvasta 3 voi havaita, että kaikilla koulutukseen valituista on aiempi tutkinto. Hakijoista yli puolella on aiempi ammattikorkeatutkinto tai yliopistotutkinto, mikä viittaa jo olemassa oleviin analyyttisiin taitoihin.

Opiskelijoita pyydettiin tutkimaan tutkinnon perusteita ja arviomaan itse, kuinka paljon heillä on aiempaa kokemusta koulutuksen perustana olevan tutkinnon aiheista. Oman arvionsa mukaan noin puolella ei ole lainkaan tai on jonkin verran aiempaa kokemusta, kun taas noin puolella on yli vuoden, joillain jopa yli 10 vuoden kokemus (kuva 4).



Kuva 4. Oma arvio aiemmasta osaamisesta vuosissa.

Koulutukseen valittujen opiskelijoiden ikäjakauma on 30-62-vuotta, kuten ilmenee kuvasta 5.



Kuva 5. Opiskelijoiden ikä ryhmiteltynä kymmenen vuoden välein.

Edellä esitetyistä lähtötiedoista voi havaita, että opiskelijoilla oli jo ennen koulutuksen alkua laajat pohjatiedot sekä iän ja aiemman työhistorian tuomaa kokemusta.

Suositus on, että ammattitutkintoa suorittavalla henkilöllä olisi vähintään kolme vuotta alan työkokemusta perustutkinnon jälkeen. Ammattitutkinnon voi suorittaa

myös ilman perustutkintoa, jos osaamista on kertynyt muulla tavalla, esimerkiksi alan töissä. Edistyneen data-analytiikan osaaja -koulutukseen haettaessa otettiin huomioon myös harrastuneisuus, motivaatio ja opiskelutausta, jos työkokemus alalta ei riittänyt.

5 Koulutuksen toteutus (toteutus)

Seitsemän kuukautta kestävä koulutus jaetaan valmistavaan koulutukseen, joka kestää noin 5 kuukautta sekä kahden kuukauden työssäoppimisjaksoon, jonka jälkeen tutkinto suoritetaan arvioimalla osaaminen suoritettun näytön perusteella.

Valmistava viiden kuukauden koulutus taas jaetaan neljään osaan:

1. ensimmäinen osa: johdanto
2. toinen osa: koneoppiminen, ohjattu ja ohjaamaton oppiminen
3. kolmas osa: lisää työkaluja dataosaajalle: NLP, neuroverkot, tekoäly ja robotiikka, versionhallinta sekä tietovarastoyhteydet
4. neljäs osa: pilvipalveluiden perusteet, Testaus ja automatisointi, MLOps, CI/CD ja Azure.

Jaotteluun vaikutti se, että osista saatiin ajallisesti saman pituisia. Samantyylliset teemat pyrittiin laittamaan samaan osaan.

Jaottelun tavoitteena on koulutuksen alussa varmistaa, että kaikilla on toimiva työympäristö ja tarvittavat työkalut käytössä. Koulutus etenee niin, että aluksi opitaan konkreettisempia ohjelmoinnin ja visualisoinnin työkaluja.

Koneoppimismalleissa käydään ensin perusteet ohjatusta oppimisesta ja sen jälkeen ohjaamattomasta oppimisesta. Näiden mallien hahmottamisessa auttaa tilastotieteen ja todennäköisyyslaskennan perusteiden hallinta. Mallintamista jatketaan kielimalleihin ja koulutuksessa käydään *sentiment*-analyysiin perusteet. Mallinnuksessa käydään lopuksi läpi neuroverkkoja, mutta tässä toteutuksessa ne on suunniteltu kolmanteen osioon, jossa otetaan lisää konkreettisia työkaluja

käyttöön, kuten Git-versionhallinta sekä tietovarastoyhteydet eli täsmällisemmin SQL-yhteyksien luominen Python-koodista käsin. Neljäs osa käsittää prosessin automatisointia viemällä malli pilviympäristöön ja testaamalla sen toimivuutta.

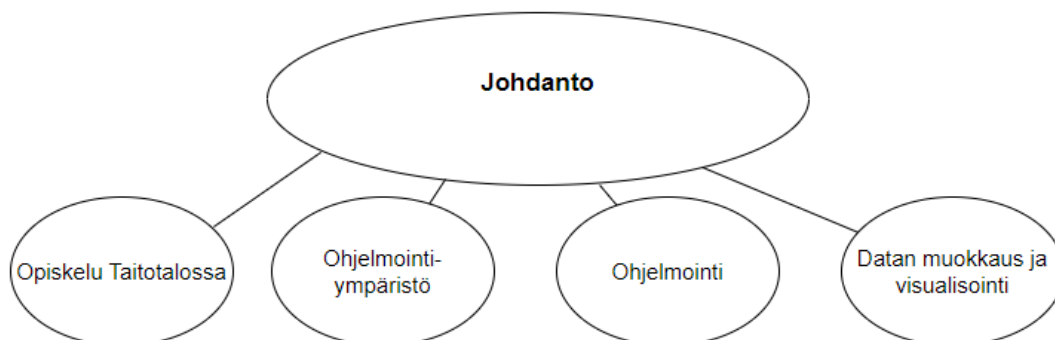
Oppimismetodeina valmistavassa koulutuksessa on luennot ja tehtävät, joita tehdään tunnilla yksittäin ja ryhmissä vaihdellen. Osa tehtävistä tuli palauttaa joko opettajalle tai yhteiseen kansioon. Useimmiten palautus tapahtuu kaikkien jaettavissa olevaan kansioon, jolloin omia vastauksia voi käydä vertaamassa toisen vastauksiin. Tämä on erityisen hyödyllistä silloin, kun ei ole olemassa yhtä absoluuttisen oikeaa vastausta, vaan samasta vastauksesta voidaan esittää erilaisia versioita. Opiskelijat toimivat ikään kuin kahdessa roolissa, toisaalta opiskelijoina, mutta toisaalta myös oman alansa asiantuntijoina. Opiskelijoilla oli käytössään kanava hyödyllisten ja aiheeseen liittyvien linkkien ja vinkkien jakoon. Jo suunnitteluvaiheessa oli tiedostettu aiheen intensiivinen ja tiivis luonne ja se, että asioiden omaksuminen vie aikaa. Ohjelma on suunniteltu niin, että kontaktiopetusta on noin kolme kertaa viikossa ja itseopiskeluaikaa keskimäärin kaksi päivää viikossa. Itseopiskelupäivät ovat olennainen osa kokonaisuutta, jotta opiskelija voi vahvistaa ja kerrata opittavia aiheita tai jatkaa eteenpäin verkkokursseja.

Työvoimakoulutuksen tutkintoon johtaviin koulutuksiin kuuluu Taitotalosta työelämään -kurssi, jonka kesto on yhteensä viisi päivää. Kurssin päivät sijoittuvat valmistavaan koulutukseen. Rajaan kurssin käsittelyn tästä insinööriyöstä pois, koska sen tuotteistaminen ja toteutus on erillinen aihe.

5.1 Ensimmäinen osa: Johdanto

Edistyneen data-analytiikan osaajakoulutuksen johdanto-osuuteen valittiin aiheita ja työkaluja, jotka orientoivat koulutuksen seuraaviin osuuksiin. Johdanto-osuuden työkalut mahdollistavat sellaisenaan datan muokkauksen, siivouksen

ja analysoinnin.



Kuva 6. Ensimmäinen osa: johdanto opetettavine aiheineen.

Kuvassa 6 esitetään ensimmäisen osuuden teemojen jako neljään osaan: opiskelu taitotalossa, ohjelmointiympäristö, ohjelmointi ja datan muokkaus ja visualisointi.

5.1.1 Johdanto-osuuden tavoitteet

Ensimmäisen jakson tavoitteena on tutustuttaa opiskelija opiskeluympäristöön, ryhmäytyä, laatia henkilökohtaiset opintosuunnitelmat (HOKS) ja saada kaikille toimiva työympäristö ja opetella ottamaan erilaisia internetissä olevia materiaaleja käyttöön.

Lisäksi tavoitteena on, että opiskelija oppii data-analytiikan perusteita eli hahmottaa kyseessä olevan ongelman, osaa hakea siihen tietoa eri tavoin, osaa muokata tietoa, osaa visualisoida sitä ja raportoida. Näiden tavoitteiden saavuttamiseksi, opiskelija

- tutustuu opiskeluun Taitotalossa
- ymmärtää ohjelmoinnin perusteet ja Python-ohjelmointikielen syntaksin
- hakee tietoa verkkosivuilta html-kielellä
- ottaa yhteyttä api-rajapintaan ja käsittelee tietoa json-formaatissa
- ottaa ohjelmointi- ja virtuaaliympäristöt käyttöön

- käsittelee dataa ohjelmallisesti Python-ohjelmointikielellä
- visualisoi dataa Python-kirjastoilla ja Power BI:llä.

5.1.2 Johdanto-osuuden toteutuksen kuvaus

Johdanto-osuuden ensimmäisellä viikolla järjestettiin aloitusinfo ja opiskelijapalveluiden info sekä aikuinen oppijana orientoimaan opiskeluun Taitotalossa sekä tulevan koulutuksen aiheisiin. Varsinaiset taitoaineet jaettiin seuraavaksi esiteltäviin kokonaisuuksiin.

Ohjelmointi ja ympäristöt

Python-ohjelmoinnin perusteiden tarkoituksena on tarjota lyhyt orientoiva alustus Python-ohjelmointiin. Tavoitteena on myös, että kaikki ymmärtävät ohjelmoinnin peruskäsitteitä, kuten muuttujat, funktiot, ehtorakenteet ja toistorakenteet, mutta toisaalta myös Pythonin tietorakenteet kuten lista, sanakirja tai array. Python on syntaksiltaan selkeä ja yksinkertainen. Tukena on laajasti erilaisia kirjastoja eri tarkoituksiin. Python on yksi kielistä, jota käytetään koneoppimisessa ja siksi oleellinen osa tätä koulutusta. [11.]

Jos tietoa käsitellään, sitä on opittava myös hakemaan. Yksi tapa hakea sitä verkosta haravoimalla html-tiedostoja tai hakemalla tarjottuja apeja. Web Scraping eli verkkoharavointi on menetelmä, jolla internetistä kerätään html-muodossa olevaa dataa. Yksi suosituimmista kirjastoista on Beautiful Soup [12] ja Selenium [13]. Apit [14] toimivat rajapintoina eri järjestelmien välillä. Esimerkiksi json-muotoista dataa voi käsitellä eri ohjelmointikielien avulla. Haravoinnin ja apien ero siinä, että apit ovat palveluntarjoajan tarjoamassa muodossa, kun haravoimalla voi hakea kaikkea tietoa, mitä verkko pitää sisällään html-muodossa.

Visual Studio Code on Microsoftin omistama, mutta ilmainen koodieditori, jolla voi koodata eri ohjelmointikieliä. VS Codeen on integroitu terminaali ja Git. Moderniin ohjelmointiin kuuluu koodieditorien käyttö ja koulutukseen VS Code valittiin siksi,

että se on ilmainen ja ladattavissa sekä Microsoft Storesta että ohjelman omilta sivuilta. Visual Studio Codeen voi ladata Jupyter Notebook -laajennuksen. [15.]

Jupyter Notebook on tiedosto, joka voidaan jakaa koodiblokkeihin ja markdown-tekstiosuuksiin. Tekstiosuuksien avulla koodia voidaan selittää ja taustoittaa. Tiedosto voidaan ajaa osissa tai kokonaan. Jupyter Notebookia hyödynnetään koneoppisessa erityisesti tutkivassa ja kokeilevassa vaiheessa. Jupyter Notebookit voidaan integroida eri järjestelmiin kuten Azure, Google Colab, Anaconda. [16.]

Anaconda on ilmainen avoimen lähdekoodin jakelualusta, jonka asennuksen mukana tulee automaattisesti yli 200 pakettia datan käsittelyyn ja koneoppimiseen. Se tarjoaa myös virtuaaliympäristön koodin ajamiseen. [17.]

Datan kerääminen, muokkaaminen, siivoaminen ja visualisointi

Useimmat koneoppimisen kirjastot vaativat, että data taulukoidaan ja siivotaan analysoitavaan muotoon. Pandas-kirjaston avulla tehdään datan muokkausta. Esimerkiksi voidaan muokata datatyyppejä, siivota, käsitellä puuttuvia tai virheellisiä arvoja ja analysoida tietoa. Pandas-kirjaston avulla voidaan tarkastella tilastollisia arvoja, kuten keskiarvoa tai minimi- tai maksimiarvoa. Pandas-kirjastoa hyödyntäen tiedot voidaan lukea tiedostosta tai tietokannasta esimerkiksi Jupyter Notebookiin ja tiedoista voi muotoilla taulukko. [18.]

NumPy-kirjastoa käytetään tieteelliseen laskentaan ja esimerkiksi moniulotteisten taulukkojen hallintaan. Tässä osuudessa NumPyä käytettiin muokkaamaan pandas-taulukoita visualisointikirjastoille luettavaan muotoon eli tukemaan visualisointeja. [19.]

Pythonilla on useita visualisointikirjastoja. Sen jälkeen, kun data on saatu siivotua, siitä voidaan hakea tietoa visuaalisin keinoin. Visualisoinnin avulla luodaan kaavioita, esimerkiksi pylväitä, ympyröitä ja taulukoita. Graafien ja kaavioiden avulla voidaan helpottaa tiedon omaksumista ja nostaa esiin olennaisia asioita.

Koulutuksessa visualisointeihin käytettiin Matplotlib- [20] ja Seaborn-kirjastoja [21].

Koulutuksessa käytetyt Python-kirjastot ovat vapaasti ladattavia avoimen lähdekoodin kirjastoja, joita ylläpidetään GitHubissa ja joilla on omat ohjaavat dokumentaationsivut. Linkit dokumentaatioon löytyvät lähdeluettelosta.

Visualisoinnissa haluttiin ottaa käyttöön myös työkalu, joka ei vaadi ohjelmointiosaamista. Olemassa olevista valittiin Power BI. Power BI on Microsoftin kaupallinen työkalu, jonka laajempi käyttö vaatii maksullisen lisenssin. Power BI:tä voidaan käyttää yhdessä muiden M365-tuotteiden kanssa. Power BI:ssä voi yhdistää dataa monista tietolähteistä. Power BI:hin liitetyn Power Query -editorin avulla dataa voi muokata. [22.]

Lisäksi opiskelijoille oli tarjolla seuraavia vapaasti valittavia opintoja itseopiskelupäiville ja opetuksen tueksi:

- Python (Metropolia 3 op) [23]
- Ohjelmoinnin perusteet 5 op (jatko 5 op) [24]
- Power BI -perusteet (Eduhouse/Metropolia) [23,25]
- Elements of AI: tekoälyn perusteet (2 op) [26]
- Excel-perusteet (Eduhouse/Metropolia) [23, 27].

Opiskelijoille annettiin mahdollisuus itse valita opinnot itseopiskelupäiville omien tarpeidensa mukaan. Opintopisteet kursseista rekisteröidään aina kuhunkin opiskelutapaan.

5.1.3 Johdanto-osuuden opiskelijapalautteet

Kysymykseen miten hyödyllisenä koet itsellesi oppimisen ja työllistymisen kannalta asteikolla 1-5, opiskelijat vastasivat keskimäärin kuvassa 7 esitetyllä tavalla. Vastaaajia oli yhteensä 17.

Kuinka hyödyllisenä näet itsellesi, jakso 1	Keskiarvo
Datan käsittely ja visualisointi (pandas, numpy, matplotlib ja seaborn) 3 pv	4,53
Johdanto koneoppimiseen, raportointimalli, pilviteknologiat	4,24
Power BI ja python	4,12
Python kertaus 2 pv	3,88
Web scraping 1 pv	3,47
Opiskelijapalvelujen info ja aikuinen oppijana 2h	3,00
Yhteensä	3,87

Kuva 7. Kysymys: kuinka hyödyllisenä näet itsellesi asteikolla 1–5, jakso 1.

Kysymykseen miten voin hyödyntää opittua omassa työssäni ja tulevaisuudessa, tuli muun muassa seuraavanlaisia vastauksia:

- toive hyödyntää opittua tulevissa työtehtävissä
- datan hyödyntäminen itsenäisesti omissa projekteissa
- reaalielämän tapahtumien mallintaminen
- soveltamisala on rajaton
- kaikki jaksolla opittu todennäköisesti olennaista tulevassa työssä
- datan ja tekoälyn ymmärtämisen hyöty ohjelmointityössä.

Koulutus ja kouluttajat saivat seuraavanlaista palautetta:

- kova tahti opiskelussa
- enemmän selkeyttä opinto-ohjelmaan ja sisältöihin
- Power BI:lle lisää päiviä
- sisältö ja opetus motivoivia
- toive resurssipankista
- syvemmin data-analytiikkaa
- kiinnostava aihe
- tsemppitoivotuksia kouluttajille.

Vastaukset on mukailtu opiskelijoiden vastauksista. Suoria lainauksia ei kooneissa ole. Palautteen perusteella luotiin alustava opetussuunnitelman koonti opiskelijoille. Johtopäätöksissä on analysoitu palautteita yleisellä tasolla.

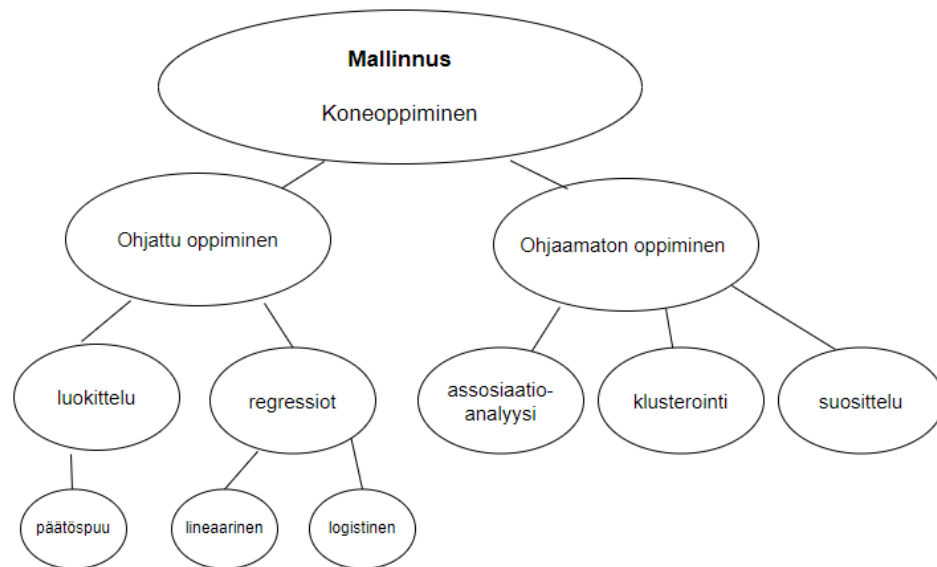
5.2 Toinen osa: Koneoppiminen, ohjattu ja ohjaamaton oppiminen

Koneoppiminen ja tekoäly ovat keskeisiä työkaluja datan analysoinnissa ja päätöksenteossa. Jakson aikana opiskelijan tulee oppia koneoppimisen mallinnuksen peruskäsitteitä sekä oppia tarkastelemaan tärkeimpiä tunnuslukuja sekä arvioida mallin hyvyttä ja mallin tuottaman tiedon oikeellisuutta.

Koneoppimisen perusteet voi jakaa ohjattuun ja ohjaamattomaan oppimiseen. Ohjatussa oppimisessa mallin opettaminen perustuu valmiiseen opetusjoukkoon, jossa opetusjoukon selittävien tekijöiden lopputulos eli vastemuuttuja on tiedossa. Mallintamisen avulla pyritään ennustamaan vastemuuttuja vastaavasta uudesta datasta. Esimerkkejä ohjatusta oppimisesta on esimerkiksi asiakasluokittelu: voiko pankki myöntää lainan. Vastaus on binäärinen kyllä tai ei. Asuntojen hinnan kehitystä ennustetaan vastaavasti lineaarisesti ja liukulukuna ja usein jollain tietyllä välillä.

Ohjaamattomassa oppimisessä ei ole erillistä opetusjoukkoa, vastemuuttujaa eikä vastemuuttujan oikeita vastauksia, vaan tavoitteena on löytää datassa piileviä rakenteita ja suhteita. Raja ei ole selkeä, vaan sisältää enemmän tulkintaa ja vaihtelua. Esimerkkeinä on assosiaatioanalyysi, jonka avulla pyritään tunnistamaan sääntöjä ja yhteyksiä, esimerkiksi jos asiakas ostaa hammasharjan niin kuinka todennäköisesti hän ostaa myös hammastahnaa sekä asiakassegmentointia kulutuskäyttäytymisen pohjalta.

Koneoppimisen kolmas laji vahvistusoppiminen jätettiin tässä koulutuksessa pois, koska sen opiskeluun ei koettu olevan ajallista, eikä taidollista resurssia.



Kuva 8. Toinen osa: Mallinnus ja koneoppiminen.

Kuvassa 8 on esitelty koulutuksessa esitellyt työkalut datan mallinnukseen. Kyseiset työkalut löytyvät pääosin avoimen lähdekoodin sci-kit learn -koneoppimisen kirjastosta. [28.]

5.2.1 Ohjatun ja ohjaamattoman koneoppimisen tavoitteet koulutuksessa

Tavoitteena on, että opiskelija tunnistaa erilaisia lähestymistapoja ratkaista jokin tietty datalla ratkaistava ongelma koneoppimisen keinoin. Tavoitteena on, että opiskelija osaa arvioida ei mallien sopivuutta ratkaisuna käsillä olevaan ongelmaan, oppii syöttämään dataa halutussa muodossa kullekin mallille, oppii arvioimaan mallin hyvyttä erilaisten kullekin mallille sopivien mittarien kautta ja arvioimaan mallin tuottaman vastauksen oikeellisuutta. Tavoitteena on myös oppia koneoppimisen peruskäsitteitä.

5.2.2 Ohjatun ja ohjaamattoman koneoppimisen toteutus

Koneoppimisen osuus jaetaan koulutuksessa kahteen ryhmään: ohjattuun oppimiseen ja ohjaamattomaan oppimiseen. Kaikki ohjatun oppimisen mallit ovat tilastollisia malleja.

Ohjattu oppiminen

Ohjatussa oppimisessa käsiteltiin luokittelua päätöspuuesimerkin avulla. Koneoppimisessa luokittelulla tarkoitetaan sitä, että malli opetetaan tunnistamaan luokkia tai kategorioita datasta valmiin opetusdatan avulla. Yksi esimerkki on roskapostin tunnistaminen normaalista sähköpostista tiettyjen kriteerien mukaan. Näin roskapostin käsittelyä voidaan automatisoida. Jossain tapauksissa luokkia voi olla enemmän kuin kaksi ja silloin puhutaan moniluokittelijasta. Päätöspuu tuottaa päätöksenteosta ja siihen liittyvistä säännöistä visuaalisen puumaisen rakenteen. Selittävillä muuttujilla ei tarvitse olla lineaarista yhteyttä.

Ohjattua oppimista voidaan lähestyä myös lineaarisen regression keinoin, jolloin vastaus ei ole luokitteleva joko tai vaan liukuluku jossain tietyssä lukuavaruudessa. Esimerkiksi taloyhtiön energian kulutusta voidaan ennustaa lämpötilan muuttuessa. Lineaarista regressiota kuvataan viivagraafilla. Lineaarisen regression vastemuuttuja ei saa liikaa korreloida selittävien muuttujien kanssa, muuten mallin selittävyys on itsestään selvää.

Kolmas tähän koulutukseen valittu ohjatun koneoppimisen malli on logistinen regressio. Logistista regressiota käytetään ennustamaan tapahtuman todennäköisyys, esimerkiksi sairastuuko henkilö uudelleen perustuen hänestä kerättyihin tietoihin.

Olennaista kaikissa koneoppimisen malleissa on validoida mallin suoritus validointi ja testijoukolla sekä erilaisilla matemaattisilla mittareilla. Kaikkien edellä mainittujen mallien tarkempaa kuvausta voi tarkastella sci-kit learn -sivuilla. [28.]

Ohjaamaton oppiminen

Ohjaamattomassa oppimisessä ei ole tarjolla oikeita vastauksia, vaan käsillä olevasta datasta pyritään löytämään rakenteita ja yhdistäviä tekijöitä, joiden avulla tuotteita voi esimerkiksi ryhmitellä kaupassa edistämään myyntiä.

Koulutukseen valittiin assosiaatioanalyysi, jossa käytettiin apriori-algoritmia. Toisin sanoen, jos ostaa olutta, niin ostaako esimerkiksi pähkinöitä tai juustoa sen lisäksi. Algoritmi analysoi ostotapahtumia ja pyrkii löytämään niistä yhdenmukaisuuksia. Mallin hyvyttä analysoitaessa arvioidaan, millaisen tuen ja luottamuksen mallin ehdottamat tuoteparit saavat. [29.]

Suosittelujärjestelmässä käytettiin avoimen lähdekoodikirjaston surprisen KNN-Basic-algoritmia, jonka avulla pyritään löytämään lähimpiä naapureita suositteluvuorossa, esimerkiksi asiakas ja hänen antamansa arvio, löytämään tälle pisteelle muita mahdollisen lähellä olevia naapureita. Näin esimerkiksi hotellimatkustajalle voidaan suositella hotellia aiempien arvioiden perusteella. [30.]

Klusterointi on asiakassegmentointia liittyen asiakkaaseen liittyviin ominaisuuksiin niin, että samankaltaiset havainnot päätyvät samaan klusteriin. Klusteroinnista käytetään joskus suomenkielistä termiä ryvästyminen. Klusteroinnissa käytettiin K-Means -algoritmia, jonka käytön esimerkkejä ja dokumentaatiota löytyy sci-kit learn -kirjastosta [28]. Algoritmin avulla voidaan määrittää eri klusterien määrän ja havainnot voidaan jakaa eri pisteiden välisen etäisyyksien mukaan. Malli alkaa laskea klustereiden keskipisteitä ja laskenta toistuu niin monta kertaa, että keskipisteiden sijainti ei oleellisesti enää muutu.

Lisäksi opiskelijoille oli tarjolla seuraavia vapaasti valittavia opintoa itseopiskelupäiville ja opetuksen tueksi:

- Artificial intelligence with Python (Metropolia 3 op)
- Introduction to Artificial Intelligence (Metropolia, 3 op)
- Todennäköisyys ja tilastomatematiikka (Metropolia, 3 op). [23.]

5.2.3 Palautteet

Kysymykseen miten hyödyllisenä koet itsellesi oppimisen ja työllistymisen kannalta asteikolla 1-5, opiskelijat vastasivat keskimäärin kuvassa 9 esitetyllä tavalla. Vastaajia oli yhteensä 13.

Kuinka hyödyllisenä näet itsellesi, jakso 2	Keskiarvo
Suosittelu (KNN) ja klusterointi (K-means), 3 pv	3,92
Tiedonhankinta API-rajapintoja käyttäen sekä Web Scraping harjoituksia , 1 pv	3,92
Assosiaatioanalyysi (apriori), 2 pv	3,85
Luokittelu ja päätöspuu 2 pv	3,85
Lineaarinen regressio, 3 pv	3,77
Logistinen regressio, 3 pv	3,62
Yhteensä	3,82

Kuva 9. Kysymys: kuinka hyödyllisenä näet itsellesi asteikolla 1–5, jakso 2.

Kysymykseen miten voin hyödyntää opittua omassa työssäni ja tulevaisuudessa, tuli muun muassa seuraavanlaisia vastauksia:

- On mielenkiintoista ennustaa asiakkaan ostokäyttäytymistä.
- Koetaan, että on opittu käyttämään eri malleja.
- Koetaan, että koneoppimisen perusteet opittu.
- On hyödyllistä kerrata aiemmin opittuja asioita.

Koulutus ja kouluttaja saivat seuraavanlaista palautetta:

- Opiskelijat olivat innostuneita ennustavasta analytiikasta ja sen sovellusalueista, mutta osa koki oppimisprosessin haasteellisena.
- Kouluttajan opetuksen motivoiva tyyli saivat kiitosta, mutta toisaalta koulutus oli nopeasti etenevä ja vaikea seurata.
- Itseopiskeluun toivottiin enemmän aikaa ja resursseja. Toivottiin myös, että opetusmateriaalit olisi jaettu hyvissä ajoin ennen koulutusta, jotta opiskelijat olisivat voineet valmistautua paremmin. Lisäksi toivottiin konkreettisia esimerkkejä siitä, miten oppimia asioita voi soveltaa työelämässä.
- Koulutuksen pilottiluonne herätti sekä myönteisiä että kriittisiä huomioita. Toisaalta opiskelijat arvostivat sitä, että he pääsivät

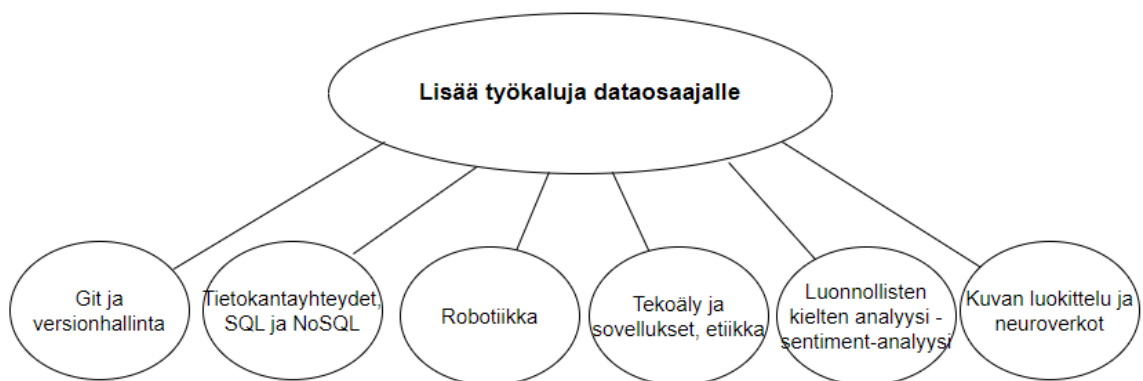
osallistumaan uudenlaiseen ja innovatiiviseen koulutukseen, toisaalta he kokivat, että koulutuksessa oli epäselvyyksiä ja sekaannuksia aiheiden käsittelyssä ja tehtävänäöissa. Tämä osoittaa, että koulutuksen suunnittelussa ja toteutuksessa on vielä kehitettävää.

- Opiskelijat pitivät hyvänä asiana sitä, että he saivat palauttaa tehtävänsä yhteiseen kansioon, josta he pystyivät näkemään ja oppimaan myös muiden töistä. Keskusteleva oppimisympäristö koettiin myös positiivisena tekijänä, joka edisti oppimista ja vuorovaikutusta.

5.3 Kolmas osa: Lisää työkaluja dataosaajalle: NLP, neuroverkot, tekoäly ja robotiikka, Git ja GitHub sekä tietovarastoyhteydet

Kolmanteen osuuteen (kuva 10) on kerätty eri aiheita liittyen dataan ja datan käsittelyyn sekä toimimiseen modernissa digitaalisessa ympäristössä. Teemansa puolesta luonnollisten kielten analyysi ja kuvan luokittelu ja neuroverkot olisivat sopineet jo edelliseen koneoppimisen jaksoon, mutta jotta jaksoista saatiin ajallisesti yhdenmittaiset, upotettiin kyseiset teemat tähän jaksoon.

Kolmannessa jaksossa Git-versionhallinta ja tietokantayhteydet sopisivat teemallisesti myös johdantopaksoon. Tekoäly ja robotiikka kertoo asioita taas ylätasolta käsin ja tekoälyssä keskitytään tällä kertaa miettimään tekoälyn ja datakäsittelyyn liittyviä eettisiä kysymyksiä. Robotiikka on yksi tekoälyn ala ja koulutuksessa ohjelmoitiin kahta robottia työpajamaisesti.



Kuva 10. Kolmas osa: Lisää työkaluja dataosaajalle.

5.3.1 Kolmannen dataosaajan työkalut osan tavoitteet

Kolmannen osuuden tavoitteena on, että opiskelija osaa data-alan tulevana ammattilaisena hyödyntää tietokantoja ja ymmärtää relaatio- ja skeematietokantojen käytön ja hyödyt datan käsittelyssä. Gitin ja versionhallinnan myötä opiskelija voi jakaa ja hallita koodia. Git on myös jatkuvan integraation oleellinen osa. Versionhallinnan avulla esimerkiksi virheiden paikannus koodissa ja sitä kautta laadunvarmistus paranee. Tekstianalyysin tavoitteena on antaa opiskelijalle työkaluja analysoida luonnollista kieltä ja kuvan luokittelun kautta tutustutaan perustasolla neuroverkkoihin ja syväoppimiseen, joita tarvitaan suurten datamallien analysoinnissa ja monimutkaisten datamallien rakentamisessa. Tekoäly on kattokäsite kaikelle koneoppimiselle ja robotiikka yksi sen osa-alueista. Robotiikan koulutusosuutta tukevat Taitotalon tarjoamat robotit, jotka esitellään sivulla 30.

5.3.2 Kolmannen dataosaajan työkalut osan toteutus

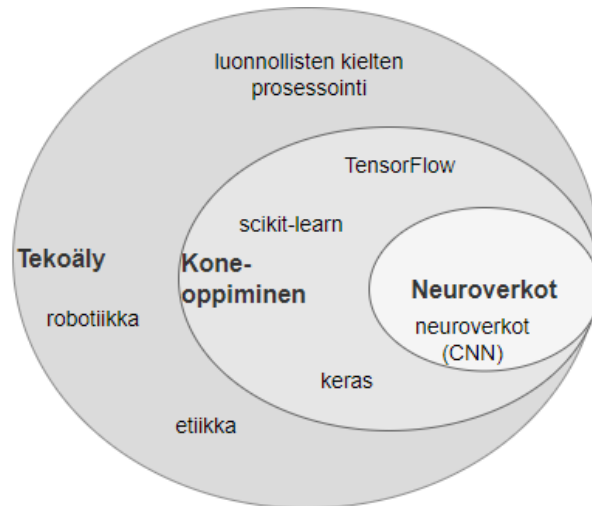
Kolmannen osan toteutus voidaan jakaa vielä kahteen osaan: ensin käsitellään yleishyödyllisiä työkaluja kaikkeen tietotekniseen osaamiseen eli versionhallinta ja relaatiotietokannat. Sitten käsitellään tekoälyä yleistasolla ja puhutaan etiikasta ja ohjelmoidaan robotteja, jotka jatkavat luonnollisen kielen prosessoinnin kautta neuroverkkoihin ja syväoppimiseen.

Tietokannat

On tärkeää, että opiskelija osaa SQL-kielen perusteet ja relaatiokantojen perusteet, jotta hän osaa käyttää relaatiotauluihin säilöttyä tietoa. Tämä tapa datan säilömiseen on laajasti käytössä. Näiden lisäksi koulutuksessa sivutaan myös skeematietokantoja, esimerkkinä MongoDB. Tietoturva datan käsittelyssä on olennaista ja tietokantojen hallinnassa on oleellista. Opiskelija oppii hyödyntämään tietokantoja Python-koodista käsin. [31;32.]

Versionhallinta

Versionhallinta on oleellinen osa ohjelmistokehitystä, myös koneoppimista, jos malliin sovelletaan jatkuvaa integraatiota ja jatkuvaa tuotantoonpanoa. Versionhallinnassa opetetaan Git:iä ja sen yhden pilviversiön GitHubin käyttöä. [33;34.]



Kuva 11. Koulutukseen valitut tekoälyn soveltamisen alat.

Tekoäly ja sen alat

Tekoäly on kattokäsite, joka kuvaa koneiden ja ohjelmistojen kykyä suorittaa tehtäviä, jotka vaativat ainakin osittain ihmisen kaltaista älykkyyttä. Laaja kattokäsite pitää sisällään niin automatisoituja päätöksentekoprosesseja kuin monimutkaisia järjestelmiä, joilla on mahdollisuus tallentaa dataa ja oppia keräämässään datasta ja muokata käyttäytymistään saadun tiedon mukaan. Tekoälyä voidaan soveltaa monella eri tavalla. Tähän koulutukseen (kuva 11) valittiin seuraavat alat: luonnollisten kielten prosessointi, neuroverkot (kuvantunnistus) ja robotiikka sekä luvussa 5.2. käsitellyt koneoppimisen mallit. Eettisten kysymysten huomioonottaminen on olennainen osa prosessia, jossa suunnitellaan laitteita, jotka pystyvät itsenäiseen toimintaan.

Astetta monimutkaisempien koneoppimismallien tutkiminen aloitetaan koulutuksessa ensin luonnollisten kielten prosessoinnista (NLP) ja kirjastona käytettiin

NLTK-kirjastoa [35]. Kirjasto auttaa jakamaan kielen osiin. Neuroverkot ovat osa syväoppimista. Se tarkoittaa, että tieto muutetaan numeeriseen muotoon ja että sitä käsitellään vähintään kolmella eri syvätasolla. Monimutkaisissa neuroverkoissa tasoja on yleensä enemmän. Koulutuksessa lähestyttiin neuroverkkoja luokitteluongelmana, jolloin tehtävänä oli luokitella kuvia perustuen mallin antamiin ennusteisiin. Tähän sovellettiin Tensorflow- ja Keras -kirjastoja. [36;37.]

Robotteja on alun perin käytetty teollisuudessa, mutta 2000-luvulta lähtien niiden palvelukäyttö ja yhteiskäyttö on monipuolistunut useille eri aloille, samoin kuin niiden määrä on kasvanut [38]. Koulutuksessa ohjelmoitiin kahta robottia. NAO6-robotille [39] ohjelmoitiin vuorovaikutteinen keskustelu. NAO6 on humanoidirobotti, jota käytetään asiakaspalvelussa, opetus- ja hoivatyössä. Humanoidirobotin tarkoitus on imitoida ihmismäistä käytöstä. Lego Mindstorms Robot Inventor Blast-robotin ohjelmoinnissa oli tavoitteena tutkia robotin antureita, valolähteitä, moottoreita ja yhteistoimintaa tietokoneen kanssa. [40.]

Lisäksi opiskelijoille annettiin ennakkotehtäväksi suorittaa SQL-perusteet (Metropolia 3 op) [23] -kurssi itseopiskelupäivinä ja tulevan opetuksen tueksi.

5.3.3 Kolmannen dataosaajan työkalut osan palautteet

Kysymykseen miten hyödyllisenä koet itsellesi oppimisen ja työllistymisen kannalta asteikolla 1-5, opiskelijat vastasivat keskimäärin kuvassa 12 esitetyllä tavalla. Vastaajia oli yhteensä 7.

Kuinka hyödyllisenä näet itsellesi, jakso 3	Keskiarvo
Git ja versionhallinta 2 pv	3,71
Luonnollisen kielen analysointia - tunneanalyysi 3pv	3,71
Tietovarastot, SQL ja NoSQL 5 pv	3,71
Tekoäly ja sovellukset, etiikka 2pv	3,57
Kuvan luokittelu 3pv	3,29
Robotiikka 2 pv	3,29
Yhteensä	3,55

Kuva 12. Kysymys: kuinka hyödyllisenä näet itsellesi asteikolla 1–5, jakso 3

Kysymykseen miten voin hyödyntää opittua omassa työssäni ja tulevaisuudessa, tuli muun muassa seuraavanlaisia vastauksia:

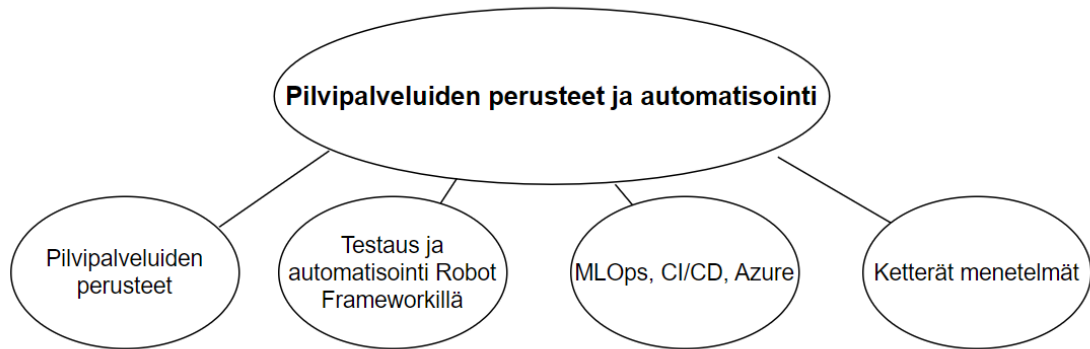
- Lähes kaikki asiat nähtiin hyödyllisenä ja käytännönläheisenä, paitsi robotiikka.
- Robotiikan tilalle ehdotettiin prosessiautomaatiota.
- Oivallettiin, että tekoälyä voi hyödyntämisestä omassa harrastuksessa.

Kolmannen osan palautteeseen vastasi seitsemän opiskelijaa. Johtopäätöksenä erityisesti tämän jakson sisältöjä ja sijoittelua koko ohjelmaan tulee seuraavassa toteutuksessa arvioida uudelleen. Robotiikkaa ei nähty hyödyllisenä työllistymisen kannalta. Havainnoivana arviona voidaan todeta, että opiskelijat alkoivat kuitenkin innokkaasti ohjelmoida robotteja ja suunnitella niille annettuja tehtäviä. Robotin ohjelmointi toimi leikki- ja luovana hetkenä kaiken ”vakavan” opiskelun välissä. Muuten palautteet koulutukselle ja kouluttajille olivat samansuuntaisia edellisten osien kanssa, joten en toista niitä tässä luvussa.

5.4 Neljäs osa: pilvipalveluiden perusteet, testaus ja automatisointi, MLOps, CI/CD

Neljännessä osassa pyritään keskittymään siihen, miten malleja ja sovelluksia voidaan saattaa tuotantoon. Useimmat sovellukset toimivat nykyään pilviympäristöissä. Rajallisen ajan vuoksi pilvipalveluissa keskitytään Microsoft-

ympäristöön, vaikka muillakin toimijoilla on vastaavia sovelluksia. Kuten kuvassa 13 huomataan, jakso on jaettu neljään osaan: 1) Pilvipalveluiden perusteet, 2) Testaus ja automatisointi, 3) MLOps, CI/CD ja Azure sekä 4) Ketterät menetelmät.



Kuva 13. Pilvipalveluiden perusteet ja automatisointi

5.4.1 Pilvipalveluiden ja automatisoinnin tavoitteet

Pilvipalvelut ovat olennainen osa IT-infrastruktuuria ja niiden peruskäsitteet on hyvä hallita, kuten tilausten ja resurssien hallinnoiminen. Testauksella voidaan varmistaa koodin toimivuus ja laatu. Tavoitteena on, että opiskelija laatii yksiker-
 taisen testaussuunnitelman ja osaa luoda yksikkötestin Pythonin unittest -kirjas-
 ton avulla [41]. Robot Frameworkilla testausta voi automatisoida [42]. Azure-ym-
 päristön opettelussa opiskelijan odotetaan osaavan automatisoida mallinnuksen
 sekä osaavan tarkastella olennaisimpia tunnuslukuja sekä arvioida mallin tuotta-
 man tiedon oikeellisuutta [43].

5.4.2 Pilvipalveluiden ja automatisoinnin toteutus

Koulutuksessa keskityttiin Azuren tarjoamiin mahdollisuuksiin pilvipalvelussa, vaikka toimijoita on muitakin [44]. Testaus on osa laadunvarmistusta ja testauk-
 sesta käytiin läpi pääperiaatteet sekä yksikkötestaus esimerkkinä. Robot Fra-
 meworkia kokeiltiin testauksen automatisointiin. Lopuksi vietiin automaattinen

mallinnus pilviympäristöön eli Azuren koneoppimisen sovellukseen. Mallin rakennuksen jälkeen tarkasteltiin tärkeimpiä tunnuslukuja ja arvioitiin mallin hyvyttä sekä mallin tuottaman tiedon oikeellisuutta [43]. Ketterät menetelmät ovat osa nykyajan projektinhallintoa ja osa CI/CD-putkea ja niitä opiskeltiin teoriatasolla yhden päivän ajan.

Lisäksi opiskelijoille oli tarjolla seuraavia vapaasti valittavia opintoja itseopiskelupäiville ja opetuksen tueksi:

- Microsoft Azure Fundamentals (Metropolia, 3 op)
- MS Project -ohjelmistot (Metropolia 5 op). [23.]

5.4.3 Pilvipalveluiden ja automatisoinnin palautteet

Kysymykseen miten hyödyllisenä koet itsellesi oppimisen ja työllistymisen kannalta asteikolla 1-5, opiskelijat vastasivat keskimäärin kuvassa 14 esitetyllä tavalla. Vastaaajia oli yhteensä 10.

Kuinka hyödyllisenä näet itsellesi, jakso 4	Keskiarvo
MLOps CI/CD Azure	4,20
Pilvipalveluiden perusteet 1 pv	3,50
Testaus 1 pv	3,20
Ketterät menetelmät	3,10
Robot Framework 1 pv	3,10
Yhteensä	3,42

Kuva 14. Kysymys: kuinka hyödyllisenä näet itsellesi asteikolla 1–5, jakso 4

Kysymykseen miten voin hyödyntää opittua omassa työssäni ja tulevaisuudessa, tuli muun muassa seuraavanlaisia vastauksia:

- Aiheet olivat kiinnostavia ja hyödyllisiä, mutta melkein kaikkia aiheita käsiteltiin vain yhden päivän verran, joten käsittely jäi pintapuoliseksi ja teoreettiseksi.

- Pilvipalveluiden osalta toivotaan AWS:n pilviympäristöön tutustumista, onhan AWS tällä hetkellä suurin toimija alalla.
- Robot Frameworkin tilalle voisi ottaa muita automatisointityökaluja, esimerkiksi Power Platformin.
- MLOps ja Azure oli yksi koko kurssin hyödyllisimmistä osista ja auttoi kertamaan jo aiemmin opittuja aiheita.

Yhteenvetona voidaan todeta, että Robot Frameworkin hyödyntäminen automaatisaatioissa ei ehtinyt saada ansaitsemaansa käsittelyaikaa toteutuksessa. Voi olla kuitenkin hyödyllisempää jatkossa keskittyä muihin automaatiotyökaluihin.

5.5 Työssäoppiminen, näyttö ja tutkinnon suorittaminen

5.5.1 Työssäoppimisjakson tavoitteet

Jakson tavoitteena on käyttää valmistavassa koulutuksessa opittuja asioita ja syvennetään opittua. Ensisijainen toive on, että opiskelija pääsee yritykseen tai vastaavaan oppimaan, miten samoja asioita toteutetaan aidossa työelämäympäristössä. Jos tämä ei ole mahdollista, tehdään Taitotalossa itsenäisesti joko ryhmässä tai yksin soveltava projekti. Jakso päättyy näytön arviointiin ja TYPO-koulutettavien tapauksessa osatutkinnon loppuun saattamiseen.

5.5.2 Työssäoppimisjakson toteutus

Opiskelijoista puolet eli 11 ei löytänyt toivomaansa työssäoppimispaikkaa ja tekivät Taitotalossa oman projektinsa. Heille järjestettiin ohjausta Taitotalossa kerran viikossa, muuten projektia tehtiin itsenäisesti. Taulukossa 3 on lueteltuna heidän valitsemansa aiheensa, teknologiat ja data.

Taulukko 3. Listaus joistakin projektiaiheista havainnollistamaan projektien aiheiden laajuutta. Lista ei ole tyhjentävä.

Aihe	Data	Työkalu
Henkilöstöhallinnon vuosiraportin laadinnassa avustaminen	Julkisen toimijan sisäinen henkilöstöhallinnon data	Power BI
Taloyhtiön energiankulutuksen ennustaminen	Taloyhtiön energiankulutustietoja	Lineaarinen regressio
Kaggle-kilpailuun osallistuminen	Kaglessa valmis data	Luonnollisten kielten analyysi koneoppimisen keinoin
Vesistöjen levänkasvuun vaikuttavat tekijät	Ilmatieteen laitos ja Suomen ympäristökeskus	Lineaarinen regressio
Monitoridatan siivoaminen	Sairaalan teho-osaston eri monitoreista tuleva data	Databricks
Videopelisuosittelija	Pelin tilastot netistä	Suosittelujärjestelmä
Chatbot, joka vastaa kyselyihin luonnollisella kielellä ja hakee tiedon laajasta datamassasta tiettyyn erikoisalaan liittyen.	Netistä haettu data verkkoharavoinnilla	Laaja kielimalli
Hotellin suosittelujärjestelmä	Esimerkki data Kagglesta	Suosittelu ja klusterointi

Neljä opiskelijaa aloitti harjoittelun Taitotalon ulkopuolella. Tähän mennessä kaksi on työllistynyt yritykseen, jossa teki harjoittelun. Kävin kolmella työpaikalla kerran seuraamassa harjoittelun kulkua. Yhtenä tavoitteena vierailuille oli verkostoitua alan toimijoiden kanssa ja löytää harjoittelupaikkoja seuraavien toteutuksien opiskelijoille.

Vuoden aikana edennyt laajojen kielimallien kehitys ja generisen tekoälyn käytön lisääntyminen näkyi muun muassa siinä, että näyttötöistä kolme käsittelee eri tavoin luonnollisten kielten mallintamista tai laajoja kielimalleja. Lisäksi yhdessä suosittelujärjestelmätyössä oli esimerkillisen oivaltavasti käytetty ChatGPT:tä

nopeuttamaan projektikehitystä ja viemään opiskelija tasolle, jolle opiskelija ei olisi yltänyt perinteisin keinoin eli opiskelemalla tutoriaaleja ja googlaamalla.

5.5.3 Työssäoppimisjakson palautteet

Opiskelijat hyödynsivät jaksoa opittujen asioiden kertaamiseen ja soveltamiseen. Erityisesti mainittiin Python-ohjelmointikielen käyttö ja eri koneoppisen kirjastojen hyödyntämisen parantuminen. Yksi kertoi vahvistaneensa osaamistaan api-raja-pintojen hyödyntämisestä eri sovellusten välillä. Näyttötöön koettiin antavan palautetta omasta osaamisesta ja palautteen toivottiin parantavan työllistymistä.

Työpaikkaohjaajat (4 kpl) olivat kaikki erittäin tyytyväisiä valitsemiinsa harjoittelijoihin ja heidän osaamiseensa. Myös koulutus sai ohjaajilta kiitettävää palautetta. Opiskelijat selvittivät ongelmia, joihin ei työpaikalla aiemmin ollut resursseja tai osaamista.

6 Vaikuttavuuden arviointi (testaus)

Tässä työssä tämän koulutuksen vaikuttavuutta arvioidaan numeerisilla mittareilla, jotka ovat opiskelijoiden työllistyminen koulutuksen aikana, opiskelijoiden menestyminen sertifikaattikokeissa opiskelijoiden antama palaute, sekä sanallisia mittareilla, joita ovat viimeisen päivän opiskelijapalautteet.

Käytettävissä oleva tieto vaikuttavuuden arviointiin rajoittuu koulutuksen aikana kerättyyn tietoon, lähinnä opiskelijapalautteiden pohjalta. Tässä työssä voidaan arvioida välittömiä oppimisvaikutuksia ja välittömiä työllistymisvaikutuksia, mutta käytössä ei ole työkaluja arvioida pitkän aikavälin vaikuttavuutta.

6.1 Opiskelijoiden työllistyminen koulutuksen aikana jälkeen

Koulutukseen haki 176 työtöntä työnhakijaa, joista haastatteluun valittiin 40. Hakijamäärä kertoo osittain, että aihe koetaan hyödyllisenä tai aiheellisenä.

Koulutukseen valittiin 22 opiskelijaa. Koulutuksen päätteeksi opiskelijoiden tilanne näytti seuraavalta:

Taulukko 4. Opiskelijoiden tilanne koulutuksen päätteeksi.

teki työharjoittelun yrityksessä (kaksi työllistyi harjoittelupaikkaan)*	4	18 %
työllistyi koulutuksen aikana	3	14 %
aloitti rekrykoulutuksen	1	5 %
yhteensä	8	36 %
teki työharjoittelun Taitotalossa	11	50 %
keskeytti	3	14 %
kaikki yhteensä	22	100 %

Taulukosta 4 huomataan, että puolet opiskelijoista päätyi tekemään työharjoittelua Taitotalossa. Moni kertoi lähettäneensä useita, jopa kymmeniä hakemuksia ja päässeensä haastatteluun, mutta eivät tulleet valituksi. Tämä kertoo ehkä enemmän työmarkkinatilanteesta kuin hakijasta. Olemme työllisyystilanteessa, jossa osaavien ja motivoituneiden työntekijöiden on vaikea työllistyä. Myös korkea ikä näytti korreloivan negatiivisesti työllistymisen kanssa. Harjoittelusta Taitotalossa tulee kuitenkin ottaa huomioon, että senkin voi nähdä hyödyllisenä jaksena, jonka aikana pääsee kertaamaan ja syventämään opittuja asioita. Keskeyttäneitä koulutuksessa oli kolme.

Neljä opiskelijaa sai toivomansa harjoittelupaikan, kolme työllistyi muualle ja yksi aloitti rekrykoulutuksen. Koulutuksen suorittaneista se on 8/19 eli 42 %. Lukua voidaan pitää alhaisena, kun tavoitteena on täystyöllistyminen, mutta näkisin sen yllättävän hyvänä tuloksena. Kaikki neljä työelämäohjaajaa olivat tyytyväisiä harjoitteluun ja kertoivat saaneensa lisäarvoa opiskelijan lyhytkestoisesta työpanoksesta. He ovat valmiita sopimaan harjoittelusta myös seuraavassa toteutuksessa, jos sopiva henkilö löytyy.

6.2 Opiskelijoiden suorittamat sertifiointikokeet koulutuksen aikana

Taitotalo tarjosi vuonna 2023 ilmaiseksi Certiportin IT-spesialisti [45] sekä Microsoftin [46] sertifiointikokeita opiskelijoille. Sertifiointikokeita suoritetaan IT-alalla, koska niiden arvioivan osoittavan objektiivista osaamista. Henkilösertifikaatteja vaaditaan toisinaan esimerkiksi julkisen alan hankekilpailutuksissa, koska niiden avulla pyritään takaamaan tilatun työn laadukkuus.

Edistyneen data-analytiikan osaajakoulutuksesta yhteensä 12 opiskelijaa eli yli puolet suoritti taulukossa 5 olevat sertifiointikokeet hyväksytysti. Hyväksytyjä suorituksia oli ryhmällä yhteensä 43, mikä tarkoittaa keskimäärin noin 3–4 hyväksytyä sertifiointikoetta per kokeen suorittaja.

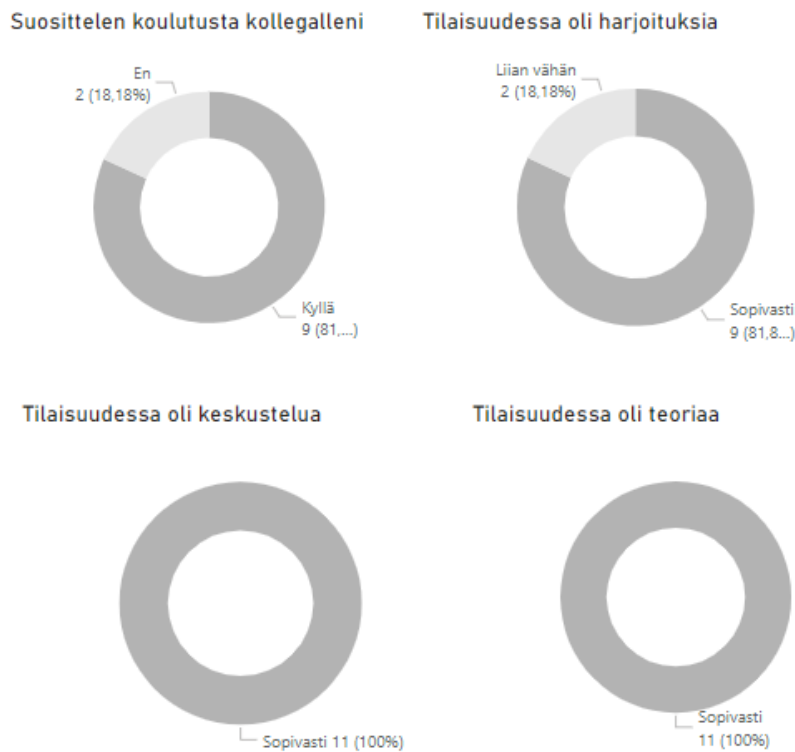
Taulukko 5. Opiskelijoiden suorittamat sertifiointikokeet.

Koe	hyväksytysti suorittaneet
AZ-900: Microsoft Azure Fundamentals	8
DP-900: Microsoft Azure Data Fundamentals	6
AI-900: Microsoft Azure AI Fundamentals	5
PL-900: Microsoft Power Platform Fundamentals	2
IT Specialist:	
Data Analytics	7
Python	5
Databases	5
Artificial Intelligence	2
Software Development	1
HTML and CSS	1
JavaScript	1
Yhteensä	43

Edistyneen data-analytiikan osaajakoulutuksen opiskelijat olivat ryhmä, josta eniten opiskelijoita osallistui Taitotalon tarjoamiin sertifiointikokeisiin. Osallistuminen kokeeseen on opiskelijoille vapaaehtoista. Siihen kannustettiin, mutta ei veloitettu. Palautteissa ei kysytty syytä poisjäämiseen.

6.3 Opiskelijoiden antama palaute

Numeerinen viimeisellä kerralla kerätty palaute (kuva 15), johon on vastannut 11 opiskelijaa, kertoo, että teoriaa ja keskustelua on ollut riittävästi, mutta että harjoituksia ja harjoittelua tarvitaan lisää.



Kuva 15. Viimeisen päivän numeerinen palaute.

Kirjallisissa päätöspalautteissa koulutus sai kiitosta ajankohtaisuudestaan, tasostaan sekä kattavasta johdannosta tekoälyyn ja koneoppimiseen. Toiveita tuli eri osa-alueista, joita ei ollut otettu tähän koulutukseen mukaan, kuten Databricks, aikasarjat ja laajat kielimallit ja Power Automation. Koulutuksen ennakkovaatimuksia voisi selventää.

6.4 Itseopiskelu

Opiskelijoille, jotka ovat työttömiä työnhakijoita, on merkittävä etu voida opiskella korkeakoulutasoisia kursseja koulutuksen yhteydessä kerryttäen opintopisteitä,

joita voi hyödyntää jatko-opinnoissa. Suoritetuista kursseista ei kerätty määrällistä dataa. Tässä on koonti, joka on luotu sanallisista palautteista:

- opiskelua eri sertifikaattikokeita varten
- Python, MOOC ja Metropolia
- Data Analytics, Helsingin yliopisto (MOOC)
- todennäköisyys ja tilastomatematiikka, 3 op, Metropolia
- Artificial intelligence with Python 3 ECTS, Metropolia
- SQL-perusteet, 3 op, Metropolia
- osallistuminen alan tapahtumiin ja työpajoihin (esim. Mimmit koodaa)
- ChatGPT, Eduhouse
- Excel 365 peruskäyttö, Eduhouse
- Microsoft Azure Fundamentals 3 ECTS, Metropolia
- Power BI, Eduhouse ja Metropolia.

7 Johtopäätökset

Koulutuksen suunnittelu ja toteutus oli kunnianhimoinen pyrkimys antaa yleiskuva datan käsittelyn ja koneoppimisen alasta seitsemässä kuukaudessa. Opiskelijat antoivat kiitettävästi palautetta kaikissa eri vaiheissa, niin sanallisesti kuin kysymyslomakkeiden kautta. Palautteista vain osa on kerätty tähän opinnäytetyöhön. Koulutuksen pilotointi sai positiivista palautetta opiskelijoilta, mutta myös paljon kehitysehdotuksia.

Käyn läpi seuraavaksi muutamia kehitysehdotuksia. Opiskelijoilta tullut palaute siitä, että asioita käsitellään liian nopeasti, eikä opetettavia aiheita ollut jäsenely riittävästi, on aiheellinen. Kurssin suunnitteluun ja toteutukseen lähdettiin ketterästi työn ohessa ja tietäen, että suunnittelua tapahtuu vielä koulutuksen aikana. Seuraava toteutus on jäsenellympi, koska ensimmäisen pilottitoteutuksen aikana luotiin opetusmateriaalit ja koulutuksen rakenne. Myös tämä opinnäytetyö on auttanut jäsentämään, mitä nykyisestä suunnitelmasta kannattaa pitää ja mitä ei. Suunnitellessa oli tiedossa, että kontaktipäivät ovat tiiviitä, joten asia oli huomioitu aikatauluttamalla kontaktiopetuksille keskimäärin kolme päivää viikossa ja itseopiskelulle kaksi päivää viikossa.

Toisaalta täytyy ottaa huomioon, että koulutus kohdistuu alalle, jossa tällä hetkellä sovelluksia päivitetään ja uusia keksitään viikoittain, mikä voi johtaa siihen, että monen tietotyöläisen työnkuva ja työnhallinta muuttuu muutaman vuoden sisällä, mikä tarkoittaa, että uusia työkaluja ja vanhojen tuttuja työkalujen uusia ominaisuuksia tulee osata ottaa nopealla tahdilla haltuun.

Alaan kuuluu tietty epätietoisuus ja epävarmuus. On osattava sietää puutteellista ja kehittyvää dokumentaatiota. Aika näyttää, mitkä uusista työkaluista jäävät pysyviksi. Esimerkiksi kun vuoden 2022 lopussa päätettiin koulutukseen ottaa yhdeksi aiheeksi luonnollisten kielten prosessointi, ei vielä tiedetty, että laajoihin kielimalleihin perustuvat sovellukset tulivat yleisesti käyttöön vuoden 2023 aikana. Laajat kielimallit ovat luonnollisen kielen prosessoinnin kehittyneempiä muotoja, jotka hyödyntävät syväoppimista tuottaakseen tekstiä. Laajat kielimallit tuovat esimerkiksi tekoälysovellusten luomisen lähes kaikkien saataville, koska niitä käytettäessä ei tarvitse osata ohjelmoida.

Huolimatta kiihtyvistä kehityskaareista, joka liittyy moninaisesti tekoälyyn pohjautuviin sovelluksiin, olen seuraavassa toteutuksessa päättänyt pitäytyä koneoppimisen perusteissa. Perusteet tulee hallita, jotta voi ymmärtää monimutkaisia malleja. Opiskelijoilla on mahdollisuus tutustua uusimpiin teknologioihin omilla projekteillaan työssäoppimisjakson aikana. Samalla neljännestä osuudesta jää pois kaikki sovelluskehitykseen liittyvät osiot, jäljelle jää vain pilvipalvelut ja Azure. Näiden tilalle tulee enemmän päiviä ensimmäiseen osioon liittyen datan hakemiseen, valintaan, siivoukseen, muokkaukseen ja visualisointiin. Datan siivoaminen on kuitenkin vaihe, joka usein vie mallinnuksesta eniten aikaa.

Opiskelijavalinnoissa tulee kiinnittää erityistä huomiota motivaatioon, työskentelytaitoihin ja opiskelutaitoihin. Taustalla on hyvä olla ohjelmallista ajattelua ja analyyttisiä taitoja, isojen kokonaisuuksien hahmottamista ja abstraktien käsitteiden hallintaa.

Jatkossa toivon, että sertifiointikokeiden suorittaminen on edelleen mahdollista budjetoida tutkinnon suorittamiseen. Ne tarjoavat objektiivisen arvion henkilön

osaamisesta ja ovat erityisesti arvostettuja IT-alalla. Tavoitteena on, että jokainen opiskelija voisi suorittaa yksi tai kaksi sertifikaattia.

Jatkossa tavoitteena on tiivistää yhteistyötä työelämän kanssa. Toive tuli ilmi palautteissa. Työllistymistä edistää, jos verkostoituu jo opiskelun aikana. Lisäksi neljä työpaikkaohjaajaa, jotka tapasin vakuuttivat, että koulutuksen aiheet ovat relevantteja työelämälle ja olivat erittäin tyytyväisiä harjoittelun toteutumiseen.

Taulukossa 6 esitellään tekemäni SWOT-analyysin tulokset koulutuksen vahvuuksista, heikkouksista, mahdollisuuksista ja uhkista.

Taulukko 6. SWOT-analyysi koulutuksesta.

	POSITIIVISET	NEGATIIVISET
SISÄISET	<p><u>Vahvuudet:</u></p> <p>Uusi koulutusala.</p> <p>Vastaa datatalouden osaajien kysyntään ja työvoimatarpeisiin.</p> <p>Tutkintoperustainen ja tähtää tieto- ja viestintätekniikan ammattitutkinnon osatutkintoon datan käsittely ja koneoppiminen (mahdollisuus jatkaa koko tutkintoon)</p>	<p><u>Heikkoudet:</u></p> <p>Kunnianhimoinen ja nopeatahtinen, mikä voi aiheuttaa opiskelijoille haasteita seurata opetusta ja omaksua uusia asioita.</p> <p>Edellyttää opiskelijoilta jonkin verran aiempia taitoja ja kokemusta ohjelmoinnista, datan käsittelystä tai tilastomatematiikasta, miten varmistaa, että opetus kohtaa opiskelijan tarpeet.</p> <p>Koulutus on pilottiprojekti, joten suunnitteluun ja materiaalin tekoon oli rajallisesti aikaa ja resursseja, mikä näkyi opiskelijapalautteissa.</p>
ULKOISET	<p><u>Mahdollisuudet:</u></p> <p>Koulutus voi edistää työttömien työnhakijoiden työllistymistä ja urakehitystä datatalouden alalla, jossa on osaajien vaje.</p> <p>Koulutus voi lisätä opiskelijoiden kiinnostusta ja motivaatiota jatkaa opintojaan ja syventää osaamistaan</p>	<p><u>Uhat:</u></p> <p>Opiskelijoiden keskeyttäminen ja työllistyminen koulutuksen aikana.</p> <p>Haavoittuvuus uusien teknologioiden suhteen, onko kehityksen seuraamiseen mahdollista riittävässä määrin, riittääkö kouluttajien oma osaaminen?</p>

datan käsittelyssä ja koneoppimisessa.

Koulutus voi luoda yhteistyömahdollisuuksia työelämän toimijoiden kanssa ja verkostoitua alan asiantuntijoiden kanssa.

Koulutus voi hyödyntää uusia teknologioita ja työkaluja, jotka ovat relevantteja ja ajankohtaisia datan käsittelyn ja koneoppimisen alalla

Koulutuksen vaikuttavuutena voidaan pitää sitä, että 42 % opiskelijoista työllistyi, suoritti harjoittelun yrityksessä tai aloitti rekrykoulutuksen ja 12 opiskelijaa suoritti yhteensä 43 sertifiointikoetta. Myös muut vakuuttivat oppineensa uusia hyödyllisiä ja mielenkiintoisia asioita koulutuksen aikana kerättyjen palautteiden perusteella.

8 Yhteenveto

Työn tavoitteena oli luoda uudelle koulutukselle opetussuunnitelma Datan käsittelyn ja koneoppimisen osatutkinnon suorittamiseksi. Osatutkinto perustuu kansallisiin ePerusteisiin. Koulutuksen tulee vastata työelämän tarpeisiin tarjoten työttömille työnhakijoille ajantasaista ja hyödyllistä tietoa. Koulutus on tutkinto-perustainen ja tähtää tieto- ja viestintäteknikan ammattitutkinnon osatutkintoon datan käsittely ja koneoppiminen. Koulutuksen nimeksi tuli Edistyneen data-analytiikan osaaja.

Koulutuksen tarvetta voi perustella sillä, että datan käsittely ja analysointi ovat tärkeitä taitoja digitaalisen murroksen aikana, ja että Suomessa on datatalouden osaajien vaje.

Koulutus toteutettiin syyskuusta 2023 huhtikuuhun 2024. Koulutuksen sisältö jaettiin neljään osaan: johdanto, mallinnus, lisää työkaluja dataosaajalle ja

pilvipalvelut. Valmistava koulutus kesti viisi kuukautta ja lopussa oli kahden kuukauden työssäoppimisjakso.

Koulutus oli pilottiprojekti, josta kerättiin palautetta opiskelijoilta. Koulutuksen vaikuttavuutta arvioitiin opiskelijoiden työllistymisen, sertifiointikokeiden, näyttötutkinnon, opiskelijapalautteiden sekä suoritettujen itseopiskelukurssien perusteella. Koulutuksen suoritti loppuun 19 opiskelijaa, joista kahdeksan eli 42 % työllistyi, suoritti harjoittelun yrityksessä tai aloitti rekrykoulutuksen. Opiskelijoista 12 suoritti yhteensä 43 sertifiointikoetta. Koulutukseen osallistuneet opiskelijat sekä harjoittelupaikkojen työpaikkaohjaajat olivat pääosin tyytyväisiä koulutukseen. Kritiikkiä sai tiivis opetustahti ja ajoittainen jäsennellyn tiedon, esimerkiksi opettävien aiheiden ennakkomateriaalien puute.

Lähteet

- 1 Tieto- ja viestintätekniiikan ammattitutkinto, OPH-2639-2020
Verkkoaineisto
<https://eperusteet.opintopolku.fi/#/fi/amatillinen/6990663/tutkinnonosat>
Luettu 28.1.2024.
- 2 Tieto- ja viestintäteknikka: Ohjelmistotuotannon opintopolku, Metropolia, aloitus syksy 2019
Verkkoaineisto
<https://opinto-opas.metropolia.fi/88094/fi/67/70361/2586/2066> Luettu 4.5.2024.
- 3 Sitran Megatrendit 2017
Verkkoaineisto
<https://www.sitra.fi/uutiset/sitran-megatrendit-2017-kasittelevat-tyon-murrosta-demokratian-rapautumista-ja-talouden-ahdinkoa/> Luettu 7.4.2024.
- 4 Valtiovarainministeriö: Pilkahduksia tulevaisuuteen - digitalisaation ja robotisaation mahdollisuudet
Verkkoaineisto
https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/79260/Pilkahduksia_tulevaisuuteen.pdf?sequence=1 Luettu 7.4.2023.
- 5 Datatalouden perusteet osaksi kansalaistaitoja
Verkkoaineisto
<https://www.sitra.fi/artikkelit/datatalouden-perusteet-osaksi-kansalaistaitoja/> Luettu 7.4.2024.
- 6 Opetushallitus: Datatalousosaamisen perusteita perusopetukseen ja toiselle asteelle
Verkkoaineisto
<https://www.oph.fi/fi/digiosaaminen/datatalousosaamisen-perusteita-perusopetukseen-ja-toiselle-asteelle/mihin-datatalous%20> Luettu 7.4.2024.
- 7 Sitra: Datatalouden kehittymisen seuranta helpottuu – koostimme työkalun tarkkailemaan mittareita
Verkkoaineisto
<https://www.sitra.fi/artikkelit/datatalouden-kehittymisen-seuranta-helpottuu-koostimme-tyokalun-tarkkailemaan-mittareita/> Luettu 7.4.2024.
- 8 Sitra: Datatalouden tilannetyökuva
Verkkoaineisto
<https://app.powerbi.com/view?r=eyJrIjojNmFmMjZyZctODFjMC00MWUyLWlWl3NmItMmQ2YzYyZGY5M2lZliwidCI6ImlwMTlyMGYxLWFiOTQtNDkzNi04NmQ3LWYzYjQwZjM4ZjRiNyIsImMiOiJh9> Luettu 7.4.2024.

- 9 Helsinki Data Science Meetup: Data Science Best Practices
Verkkoaineisto
https://www.youtube.com/watch?v=iNwDkUhpXJA&ab_channel=DAIN-Studios Luettu 10.3.2024.
- 10 Koneoppimisinsinööri – taidot ja urapolku
Verkkoaineisto
<https://www.unite.ai/fi/koneoppimisinsin%C3%B6%C3%B6ritaitojen-urapolku/> Luettu 28.4.2024
- 11 Python-ohjelmointi
Verkkoaineisto
<https://www.python.org/> Luettu 7.4.2024.
- 12 Web Scraping
Verkkoaineisto
<https://beautiful-soup-4.readthedocs.io/en/latest/> Luettu 7.4.2024.
- 13 Selenium-kirjasto verkkoselainten automatisointiin
Verkkoaineisto
<https://www.selenium.dev/documentation/> Luettu 7.4.2024.
- 14 Python API Tutorial: Getting Started with APIs
Verkkoaineisto
<https://www.geeksforgeeks.org/python-api-tutorial-getting-started-with-apis/> Luettu 28.4.2024
- 15 Visual Studio CODE
Verkkoaineisto
<https://code.visualstudio.com/download%20> Luettu 21.4.2024
- 16 Jupyter Notebook
Verkkoaineisto
<https://jupyter.org/> Luettu 28.4.2024
- 17 Anaconda
Verkkoaineisto
<https://anaconda.cloud/package-categories> Luettu 21.4.2024.
- 18 pandas
Verkkoaineisto
<https://pandas.pydata.org/> Luettu 21.4.2024.
- 19 NumPy
Verkkoaineisto
<https://numpy.org/> Luettu 21.4.2024.

- 20 Matplotlib
Verkkoaineisto
<https://matplotlib.org/> Luettu 21.4.2024.
- 21 seaborn
Verkkoaineisto
<https://seaborn.pydata.org/> Luettu 21.4.2024.
- 22 Power BI
Verkkoaineisto
<https://learn.microsoft.com/fi-fi/power-bi/fundamentals/power-bi-overview>
Luettu 21.4.2024.
- 23 Tieto- ja viestintätekniiikan tutkinto-ohjelman NonStop Virtuaalitarjonta
Verkkoaineisto
<https://www.metropolia.fi/fi/opiskelu-metropoliassa/avoin-amk/tieto-ja-viestintateknikka-virtuaaliopinnot#koneoppiminen> Luettu 21.4.2024.
- 24 Ohjelmoinnin perusteet 5 op ja jatko 5 op.
Verkkoaineisto
<https://ohjelmointi-23.mooc.fi/> Luettu 21.4.2024.
- 25 Power BI Eduhouse
Verkkoaineisto
<https://app.eduhouse.fi/palvelu/koulutuskokonaisuudet/39579943-power-bi> Luettu 21.4.2024.
- 26 Elements of AI: Tekoälyn perusteet 2 op.
Verkkoaineisto
<https://www.elementsofai.com/fi/>
Luettu 21.4.2024.
- 27 Excel 365 peruskäyttö Eduhouse.
Verkkoaineisto
<https://app.eduhouse.fi/palvelu/opintopolut/75904428-excel-perusteet>
Luettu 21.4.2024.
- 28 sci-kit learn
Verkkoaineisto
<https://scikit-learn.org/stable/index.html> Luettu 27.4.2024.
- 29 apriori
Verkkoaineisto
<https://pypi.org/project/apriori/> Luettu 27.4.2024.

- 30 surprise
Verkkoaineisto
https://surprise.readthedocs.io/en/stable/prediction_algorithms.html%20
Luettu 27.4.2024.
- 31 Python MySQL
Verkkoaineisto
https://www.w3schools.com/python/python_mysql_getstarted.asp Luettu
4.5.2024
- 32 MongoDB
Verkkoaineisto
<https://www.mongodb.com/> Luettu 4.5.2024
- 33 Git
Verkkoaineisto
<https://git-scm.com/> Luettu 4.5.2024
- 34 GitHub
Verkkoaineisto
<https://github.com/> Luettu 4.5.2024
- 35 Natural Language Toolkit
Verkkoaineisto
<https://www.nltk.org/> Luettu 4.5.2024
- 36 TensorFlow
Verkkoaineisto
<https://www.tensorflow.org/> Luettu 4.5.2024
- 37 Keras
Verkkoaineisto
<https://keras.io/> Luettu 4.5.2024
- 38 Opettajien robotiikan täydennyskoulutuksen materiaalit (ei julkisia).
Verkkoaineisto
<https://www.roboticscampus.com/hankkeet/avoin-robotlab/opettajien-robotiikan-taydennyskoulutus/> Luettu 4.5.2024
- 39 NAO6-robotti
Verkkoaineisto
<https://www.aldebaran.com/en/support/nao-6/downloads-sofwares> Lu-
ettu 4.5.2024
- 40 Lego Mindstorms Robot Inventor Blast
Verkkoaineisto

<https://github.com/LEGO/MINDSTORMS-Robot-Inventor-hub-API?tab=readme-ov-file> Luettu 4.5.2024

- 41 Yksikkötestaus
Verkkoaineisto
<https://docs.python.org/3/library/unittest.html> Luettu 11.3.2024.
- 42 Robot Framework
Verkkoaineisto
<https://docs.robotframework.org/docs> Luettu 11.3.2024.
- 43 azure AI | Learning Studio
Verkkoaineisto
<https://ml.azure.com/> Luettu 11.3.2024
- 44 Microsoft Azure
Verkkoaineisto
<https://learn.microsoft.com/en-us/training/azure/> Luettu 11.3.2024
- 45 Henkilösertifikaatit
Verkkoaineisto
<https://certiport.pearsonvue.com/Certifications/ITSpecialist/Certification/Certify> Luettu 8.3.2024.
- 46 Henkilösertifikaatit
Verkkoaineisto
https://learn.microsoft.com/en-us/credentials/browse/?credential_types=certification Luettu 8.3.2024.

Kuva kansalaisdatatieteestä

