



Ennustemallien käyttömahdollisuudet julkishallinnon rajapinnoilla

Palvelutarpeen ennakointi TYM-yhteistoimintamallissa

Ammattikorkeakoulututkinnon opinnäytetyö

Tietojenkäsittelyn koulutus

Vuosi 2025

Mari Asplund

Koulutus	Tietojenkäsittelyn koulutus	
Tekijä	Mari Asplund	Vuosi 2025
Työn nimi	Ennustemallien käyttömahdollisuudet julkishallinnon rajapinnoilla	
Ohjaaja	Tommi Lahti	

Ennustemalli on koneoppimiseen perustuva malli, jonka tarkoitus on arvioida tulevia arvoja olemassa olevan datan perusteella. Opinnäytetyön tarkoituksena oli selvittää, kuinka ennustemallia olisi mahdollista hyödyntää monitoimijaisen palvelutarpeen arvioinnin tukena. TYM-yhteistoimintamalli on lakisääteinen työvoimaviranomaisen, Kelan ja hyvinvointialueen yhteinen toimintamalli. Se on tarkoitettu työttömälle asiakkaalle, jonka työllistyminen edellyttää näiden viranomaisten palvelujen yhteensovittamista ja työllistymisen tukea. Ennustemallin tavoitteena olisi tunnistaa asiakkaan tuen tarve varhaisessa vaiheessa ja ohjata hänet tarkoituksenmukaisten palvelujen piiriin. Tässä opinnäytetyössä ei keskitytä itse ennustemallin tekniseen rakentamiseen, sillä teknologista osaamista tähän on jo laajasti saatavilla. Painopisteenä on julkishallinnon toimijoiden välinen tiedonvaihto, joka on keskeinen este ennustemallien täysimittaiselle hyödyntämiselle, koska tiedon yhdistäminen ennustemallien tueksi ei ole vielä mahdollista. Työssä tarkasteltiin erityisesti tiedon jakamiseen, yhteentoimivuuteen ja sääntelyyn liittyviä edellytyksiä sekä vertailtiin kansainvälisiä esimerkkejä ennustemallien käytöstä julkishallinnossa.

Opinnäytetyön teoreettisessa osuudessa määritellään työn kannalta keskeiset käsitteet tekoälystä, ennustemalleista, TYM-yhteistoimintamallista ja julkishallinnon digitalisaatiosta. TYM-mallissa on useampi viranomainen, joita velvoittavat organisaatiokohtaiset vastuut ja säädökset, jotka on otettava huomioon, kun suunnitellaan ennustemallien käyttöä. Teoriaosassa sivutaan myös tekoälyn eettistä käyttöä ja tekoälyn käyttöä säänteleviä lakeja, jotka ovat oleellisia ennustemallien kokeilussa ja käyttöönotossa. Teoriaosa luo pohjan kirjallisuuskatsauksen ymmärtämiseen sosiaaliturvan, julkishallinnon digitalisaation sekä toiminnallisten ja oikeudellisten edellytysten näkökulmasta.

Opinnäytetyön toiminnallinen osuus on kartoittava kirjallisuuskatsaus, jonka tavoitteena on muodostaa yleiskuva ulkomailla kehitettyjen tai kehitteillä olevien ennustemallien ja muiden tunnistamistyökalujen käytöstä. Katsauksessa tarkastellaan myös maiden välisiä eroja dataekosysteemeissä, teknisissä ratkaisuissa sekä oikeudellisissa ja eettisissä haasteissa. Ulkomaisten käytäntöjen pohjalta peilataan Suomen nykytilannetta ja sen kehittämistarpeita.

Katsauksessa havaittiin, että ennustemallit voivat parantaa palvelujen kohdentamista, mutta ne vaativat pitkäaikaista kehittämistä ja suuria tiedonvaihdon infrastruktuurin muutoksia ja juridisia linjauksia. Teknologinen valmius kasvaa nopeammin kuin strategisesti ja hallinnollisen koordinoinnin osalta ollaan valmiita. Monen toimijan väliseen tiedonvaihtoon ja yhteisrekistereihin liittyy paljon epävarmuuksia, joihin ei vielä ole vastauksia. Suomen digitaalinen infrastruktuuri eroaa muiden maiden ratkaisuista, joten eri viranomaisten tiedonvaihtoa ei voida suoraan rakentaa ulkomaisilla malleilla jäljitellen. Muiden maiden malleista voidaan kuitenkin ottaa oppia ja varautua pilotoimalla malleja joustaviksi, jotta niitä voidaan mukauttaa tuleviin järjestelmiin ja ratkaisuihin.

Avainsanat Koneoppiminen, ennustemalli, julkishallinnon digitalisaatio, TYM-yhteistoimintamalli
Sivut 70 sivua ja liitteitä 1 sivua

DP Degree Programme in Business Information Technology
Author Mari Asplund Year 2025
Subject The use of predictive models at the interfaces of public administration
Supervisors Tommi Lahti

A predictive model is a machine learning–based model designed to estimate future values based on existing data. The purpose of this thesis was to examine the potential of predictive models in supporting multi-agency assessments of service needs. The TYM cooperation model is a statutory framework that brings together the public employment service, Kela (the Social Insurance Institution of Finland), and the wellbeing services counties. It serves unemployed clients whose employment requires coordinated services and targeted support from multiple authorities. Predictive model could help identify support needs at an early stage and guide clients to the most appropriate services.

The study does not focus on the technical development of predictive models, as such expertise is already widely available. Instead, it examines data sharing between public sector actors, which is currently the main barrier to the full use of predictive models, as data from different authorities cannot yet be combined. The research explores the prerequisites for data sharing, interoperability, and regulation, and compares international examples of predictive model use in the public sector.

The theoretical section defines key concepts such as artificial intelligence, predictive models, the TYM model, and public sector digitalization. In the TYM model, several authorities operate under separate legal and organizational responsibilities, which must be considered when planning the use of predictive models. The section also addresses AI ethics and relevant legislation, both essential to testing and adoption. This framework supports the literature review, which approaches the topic from the perspectives of social security, public sector digitalization, and the functional and legal conditions for implementation.

The practical section is a scoping literature review that maps the use of predictive models and other early identification tools developed or in development abroad. It examines cross-country differences in data ecosystems, technical solutions, and legal and ethical challenges, and uses these to reflect on Finland’s situation and development needs.

The review found that predictive models can improve service targeting but require long-term development, major changes to data-sharing infrastructure, and clear legal decisions. Technological capabilities are evolving faster than strategic and administrative readiness. Multi-agency data sharing and joint registers face significant uncertainties with no clear solutions yet. Finland’s digital infrastructure differs from other countries, preventing direct replication of foreign models, but valuable lessons can still be drawn. Pilot projects should be designed for flexibility so they can adapt to future systems and solutions.

Keywords Machine learning, predictive model, Digitalization of the Public Sector, TYM cooperation model
Pages 70 pages and appendices 1 page

Sisällys

1	Johdanto	1
2	Tekoäly ja koneoppiminen.....	2
2.1	Ennakoiva tekoäly.....	4
2.1.1	Ennustemalli.....	4
2.1.2	Ennustemallin kehittämisen prosessi.....	6
2.1.3	Datan rooli ennustemallin perustana.....	9
3	TYM-yhteistoimintamalli	10
3.1	TYM-mallin toimijat ja palveluprosessi	10
3.2	TYM-laki.....	12
4	Digitalisaatio julkishallinnossa	13
4.1	Dataekosysteemit ja data-avaruus	15
4.2	SOTE-tekoälyn ekosysteemi	18
4.3	Eettinen tekoälyn käyttö	18
4.4	Tekoälyn käytön keskeinen lainsäädäntö	21
5	Opinnäytetyön toteuttaminen	26
6	Ennustemallien käyttötapaukset.....	28
6.1	Käyttötapaukset Suomessa.....	28
6.2	Käyttötapaukset vertailumaissa.....	31
6.2.1	Alankomaat	31
6.2.2	Tanska.....	35
6.2.3	Viro.....	38
6.3	Käyttötapausten vertailu.....	41
6.3.1	Toimijoiden rooli ja datanjakamisen kulttuuri.....	43
6.3.2	Teknologiset ja toiminnalliset edellytykset.....	44
6.3.3	Lakien soveltaminen ja tulkinta	45
7	Vertailun tulokset.....	47
7.1	Kehittämisessä huomioitavaa.....	52
7.2	Mahdolliset riskitekijät	54
8	Johtopäätökset ja pohdinta	57
9	Yhteenveto.....	61
	Lähteet.....	63

Sanasto

AI Act	Artificial Intelligence Act eli EU:n tekoälysäädös
Data Act	EU:n datasäädös
Data-avaruus	Jaettu infrastruktuuri datan hallintaan ja käyttöön
Dataekosysteemi	Useiden toimijoiden välinen verkosto, jossa jaetaan sekä hallinnoidaan dataa
DPIA	Data protection impact assessment eli tietosuojaa koskeva vaikutusten arviointi
Ennustemalli	Koneoppimiseen perustuva menetelmä, jonka tavoitteena on ennustaa tulevia tapahtumia, käyttäytymistä tai arvoja aiempaan dataan perustuen
EUIF	Viron työvoimapalvelu ja työttömyysvakuutusrahasto
GDPR	General Data Protection Regulation eli EU:n yleinen tietosuojasetus
OECD	Taloudellisen yhteistyön ja kehityksen järjestö
Regulatory Sandbox	Sääntelyhiekkalaatikko eli säännelty ja valvottu testiympäristö uusille teknologioille
Sosiotekninen systeemi	Organisaatio tai muu toimija, jossa tarvitaan ihmistoimijoiden ja teknologisten järjestelmien sujuvaa yhteistyötä
TYM-yhteistoimintamalli	Työllistymistä edistävä monialaisen tuen yhteistoimintamalli

Kuvat – esimerkki

Kuva 1. TYM-prosessi mukaillen (KEHA-keskus, 2024)	11
Kuva 2. Datataloudessa tarvittavia kyvykkyyksiä mukaillen (Vastuu Group, 2025).....	17

Taulukot

Taulukko 1. Käyttötapausten vertailu	42
Taulukko 2. Suunnittelussa huomioitavaa	53
Taulukko 3. Mahdolliset riskitekijät	55

1 Johdanto

Opinnäytetyössä tarkastellaan ennustemallien käyttömahdollisuuksia julkishallinnossa ja sosiaaliturvassa, erityisesti työllistymistä edistävän monialaisen tuen yhteistoimintamallin näkökulmasta. Ennustemalleja on jo kokeiltu esimerkiksi työkyvyttömyysriskin ennustamiseen ja malleista on saatu lupaavia tuloksia. Jos voisimme ajoissa ja mahdollisimman tarkasti ennustaa palvelutarvetta, varhainen puuttuminen ja oikea-aikainen tarvittavien palvelujen piiriin ohjaaminen tukisi kansalaisten hyvinvointia ja säilymistä työelämässä. Tavoitteena olisi ennustemalli, joka koulutusdatan ja eri toimijoiden tietojen avulla tunnistaisi asiakkaiden työkyvyn riskejä, jotta heidät voitaisiin ohjata varhaisessa vaiheessa monitoimijaisen tuen piiriin. Tällä hetkellä toimijoiden välisten tarvittavien tietojen yhdistely ei ole mahdollista. Julkisissa palveluissa syntyy nykyisin suuria määriä tietoa, mutta tämän tiedon hyödyntämiseen liittyy merkittäviä haasteita, erityisesti tiedon yhdistämisen, jakamisen ja yhteentoimivuuden osalta.

Työllistymistä edistävän monialaisen tuen yhteistoimintamalli (TYM) on lakisääteinen toimintamalli, jossa työvoimaviranomainen, Kansaneläkelaitos ja hyvinvointialue sovittavat yhteen palveluitaan työttömien asiakkaiden tueksi. Tavoitteena on tarjota suunnitelmallista, asiakaslähtöistä ja monitoimijaista tukea niille työnhakijoille, joiden työllistyminen edellyttää useiden viranomaistahojen tiivistä yhteistyötä. TYM-yhteistoimintamallin ymmärrys on oleellista, jotta voidaan hahmottaa yhteistoimijuuden haasteet tiedonjaon näkökulmasta. Työn teoriaosassa käydään läpi TYM-yhteistoimintamallin lisäksi ennustemallin perusteita ennakoivasta tekoälystä datan rooliin, sekä julkishallinnon digitalisaation tilaa dataekosysteemeistä keskeiseen lainsäädäntöön.

Minulla on yli kymmenen vuoden työhistoria sosiaalietuuksien parissa ja tietojenkäsittelyn tutkinnon myötä olen pyrkinyt yhdistämään omaa substanssiosaamistani etuustyön tekniseen kehittämiseen. Työ sai alkunsa siitä tarpeesta, että vaikka ennustemallilla voitaisiin tunnistaa työkyvyttömyyden tai pitkäaikaisen työttömyyden riskissä olevat henkilöt ja ohjata heidät varhain monitoimijaisen tuen piiriin, tällaisen ennustemallin rakentaminen ei ole tällä hetkellä mahdollista, koska TYM-yhteistoimintamallin toimijoiden tietoja ei voida yhdistää ennustemalliin. Tässä opinnäytetyössä tarkastellaan, millaisia oikeudellisia, teknisiä ja eettisiä edellytyksiä tiedon yhdistämiseen liittyy. Työn lähtökohtana ei ole uuden ennustemallin tekninen kehittäminen, sillä teknologista osaamista tähän on jo olemassa

useissa organisaatioissa. Työssä tarkastellaan erityisesti sitä, mitä vaatimuksia ennustemallien kehittäminen yhteiseksi työkaluksi asettaa viranomaisten monialaisessa yhteistyössä. Palveluohjaus työttömille ja osatyökykyisille työnhakijoille on nykyisellään hallinnollisesti raskas ja hidas prosessi. Työn oletuksena on, että dataekosysteemi olisi hyödyllinen kaikille osapuolille, virastoille sekä asiakkaille. Erityishuomio kiinnitetään digitalisaatioon, toimijoiden väliseen tiedonjakoon ja sääntelyn asettamiin edellytyksiin sekä tarkastellaan kansainvälisiä esimerkkejä ennustemallien käytöstä julkisessa hallinnossa. Työssä on tarkoitus selvittää, kuinka julkishallinnon digitalisaatio tukee datan jakamista julkishallinnon virastojen kesken, sekä missä mennään nyt ja mikä on tulevaisuuden tavoite. Työssä tarkastellaan mitä asiakkaan tietoa voidaan käyttää ja miten mahdollistetaan tiedonkulku viranomaisten välillä, miten sama toteutuu eettisesti ja oikeudellisesti kestävästi jossain muussa EU-maassa työllisyyden tukemisessa ja työkyvyttömyyden ehkäisemisessä. Tarkastelun pääpaino on siinä, miten tutkimuskysymykset ilmenevät ja konkretisoituvat ennustemallien kehittämisessä TYM-yhteistoimintamallin kontekstissa.

Opinnäytetyön tutkimuskysymykset on rajattu seuraaviin:

1. Millaisia oikeudellisia, teknisiä ja organisatorisia edellytyksiä datan jakamiseen ja yhteentoimivuuteen liittyy julkishallinnon ennustemallien käytössä?
2. Kuinka tiedonjaon rajoitukset ja yhteisrekisterinpito vaikuttavat ennustemallien hyödyntämiseen TYM-palveluiden näkökulmasta, eli mitä dataa voidaan käyttää?
3. Miten ennustemallien käyttöä on toteutettu muissa EU-maissa työllisyyden tukemisessa ja työkyvyttömyyden ehkäisyssä, ja mitä oppeja niistä voidaan soveltaa Suomeen?
4. Miten varmistetaan ennustemallien eettisesti ja oikeudellisesti kestävä käyttö monitoimijaisessa yhteistyössä?

2 Tekoäly ja koneoppiminen

Tekoäly on tietojenkäsittelytieteen ala, jonka tavoitteena on kehittää järjestelmiä, jotka oppivat, tekevät päätöksiä ja ratkaisevat ongelmia ihmisen älykkyyden tapaan.

Tekoälyjärjestelmät vastaanottavat tietoa ympäristöstään, suorittavat laskennallisia operaatioita päätöksenteon tueksi ja toteuttavat toimia itsenäisesti asetettujen tavoitteiden

saavuttamiseksi. Näiden järjestelmien suorituskyky voi parantua ajan myötä koneoppimisen avulla, jolloin tekoäly kykenee mukautumaan uusiin tilanteisiin ja kehittämään toimintaansa kokemuksen perusteella. (Xiao, 2022, ss. 27–28)

Tekoälyohjelmistot perustuvat ihmisen suunnittelemiin algoritmeihin, mutta niitä voidaan lisäksi ohjelmoida oppimaan uutta ja itsenäisesti kehittämään suorituskykyään. Tekoälyn keskeisenä tavoitteena on tukea ihmisen toimintaa ja tehostaa työskentelyä esimerkiksi käsittelemällä suuria tietomääriä, joiden analysointi veisi ihmiseltä vuosia. AI-malli eli tekoälymalli on koneoppimiseen tai algoritmeihin pohjautuva järjestelmä, jonka avulla voidaan suorittaa ja automatisoida erilaisia tehtäviä tai ratkaista ongelmia, kuten tulevan kehityksen ennustaminen. Algoritmi tarkoittaa niiden ohjeiden joukkoa, jotka kertovat miten jokin tehtävä tai prosessi suoritetaan. Oleellisena osana tekoälyn käyttöä on termi Big data, jolla tarkoitetaan suuria datamääriä, jotka ovat liian runsaita ja monimutkaisia käsiteltäväksi ja analysoitavaksi perinteisin tekniikoin. Big dataa tuotetaan joka päivä jatkuvasti ja valtavia määriä, ja se voi olla peräisin monista eri digitaalista tietoa keräävistä ja tuottavista lähteistä. (Kallio & Kolari, 2023, Osa 1)

Koneoppimisella (engl. *machine learning*) tarkoitetaan tekoälyn osa-aluetta, jossa tietokoneet oppivat automaattisesti parantamaan suorituskykyään kokemuksen perusteella ilman että kehittäjän tarvitsee ohjelmoida yksittäisiä sääntöjä valmiiksi. Koneoppiminen voi olla ohjattua, jossa koneelle annetaan oikea vastaus, tai itsenäistä, jolloin kone päättää vastauksen itse datan perusteella. Menetelmä perustuu algoritmeihin, jotka analysoivat suuria tietomääriä, tunnistavat säännönmukaisuuksia ja käyttävät näitä malleja tulevien tapahtumien ennustamiseen tai päätösten tekemiseen. IBM:n mukaan koneoppiminen toimii erityisesti tilanteissa, joissa perinteinen ohjelmointi olisi liian monimutkaista tai aikaa vievää, ja sitä sovelletaankin laajasti muun muassa suositusjärjestelmissä, jotka tarjoavat käyttäjille personoituja ehdotuksia ja petostentorjunnassa, jossa tunnistetaan poikkeavaa käyttäytymistä esimerkiksi maksutapahtumissa. (IBM, 2021; Kallio & Kolari, 2023, Osa 1)

Syväoppiminen on yksi koneoppimismenetelmä, joka perustuu monikerroksisiin neuroverkkoihin ja kykenee käsittelemään myös rakentamatonta dataa, kuten kuvia ja tekstiä, oppien automaattisesti olennaiset piirteet suurista tietomassoista. Neuroverkot koostuvat kerroksittain järjestetyistä solmuista, jotka mallintavat ihmisaivojen toimintaa, jolloin tekoäly oppii paitsi tunnistamaan paitsi tunnistamaan asioita, niin myös tulkitsemaan kontekstia ja täydentämään puuttuvia tietoja. (IBM, 2021; Kallio & Kolari, 2023)

2.1 Ennakoiva tekoäly

Ennakoiva tekoäly hyödyntää tilastollista analyysiä ja koneoppimista mallintaakseen käyttäytymismalleja, tunnistaakseen kausaalisuhteita ja ennustaakseen tulevia tapahtumia. Organisaatiot käyttävät ennakoivaa tekoälyä muun muassa tulevien tapahtumien ennustamiseen, riskien arviointiin ja päätöksenteon tueksi. (Mucci, 2024)

Analyttikot ovat perinteisesti käyttäneet ennakoivaa analytiikkaa tehdäkseen tietoon perustuvia päätöksiä. Ennakoiva tekoäly kuitenkin nopeuttaa tätä prosessia ja voi parantaa tarkkuutta, koska koneoppimisalgoritmit pystyvät nopeasti käsittelemään valtavia määriä tietoa, jopa vuosikymmenten ajalta. Näiden analyysien avulla organisaatiot voivat valmistautua tuleviin trendeihin. Ennakoivaa tekoälyä hyödynnetään laajasti asiakaskäyttäytymisen analysoinnissa ja päätöksenteon optimoinnissa eri toimialoilla. Sillä voidaan ennustaa esimerkiksi asiakaspoistumaa, toimitusketjun häiriöitä tai mekaanisia vikoja – mahdollistaen ennakoivan suunnittelun ja luotettavat ennusteet. (Mucci, 2024)

2.1.1 Ennustemalli

Koneoppimiseen perustuvat ennustemallit voidaan yleisesti jakaa kahteen pääryhmään: luokittelumalleihin ja määrällisiin malleihin. Luokittelumallit arvioivat, kuinka todennäköisesti jokin tietty tapahtuma toteutuu, ja vastaavat yleensä kysymykseen siitä, tapahtuuko ilmiö vai ei. Luokitteluongelmassa voi olla myös useita vaihtoehtoisia lopputuloksia. Esimerkiksi voidaan arvioida, kuuluuko asiakas suuremmalla todennäköisyydellä nuorten, keski-ikäisten vai ikääntyneiden ryhmään. Määrämallit puolestaan ennustavat jonkin muuttujan suuruutta, kuten ajan kestoa, myyntimäärää tai jonkin ilmiön intensiteettiä. Luokittelumalleihin viitataan usein käsitteellä klassifikaatiomallit, kun taas määrämalleista käytetään nimitystä regressiomallit. (Kananen, Puolitaival, 2019, s. 109)

Ennustemalleihin perustuvia koneoppimiskäyttöisiä ratkaisuja hyödynnetään yhä useammin täydentämään tai jopa korvaamaan asiantuntija-arvioita ja manuaalista päätöksentekoa. Ne soveltuvat erityisesti tarkasti rajattuihin tehtäviin, joissa tarvitaan nopeaa ja kustannustehokasta analyysiä suurista tietomääristä. Kehittämisen jälkeen mallit ovat yleensä edullisia käyttää ja voivat tuottaa tarkkoja ennusteita. Mallien käyttökelpoisuus ja ennustetarkkuus on kuitenkin arvioitava tapauskohtaisesti, sillä konekin voi tehdä virheitä. Ennustemallit perustuvat tilastollisiin jakaumiin, ja vaikka ne ovat lähtökohtaisesti

puolueettomia, niiden oppima riippuu käytetystä datasta. Mikäli opetusdata sisältää ennakkoluuloja tai syrjiviä piirteitä, malli voi toistaa niitä. Siksi vastuu mallin laadusta ja eettisyydestä on aina ihmisellä. (Kananen, Puolitaival, 2019, ss. 109–111)

Ennustemallit (engl. predictive models) ovat tilastollisia tai koneoppimiseen perustuvia menetelmiä, joiden tavoitteena on ennustaa tulevia tapahtumia, käyttäytymistä tai arvoja aiempaan dataan perustuen. Mallien ytimessä on ajatus siitä, että menneistä havaintotiedoista voidaan löytää säännönmukaisuuksia, joita voidaan hyödyntää tulevaisuuden todennäköisyyksien arvioinnissa. Ennustemallit voivat perustua perinteisiin tilastollisiin menetelmiin, kuten regressioanalyysiin, tai nykyaikaisempiin koneoppimistekniikoihin, kuten päätöspuihin, tukivektorikoneisiin tai neuroverkkoihin. Mallin valintaan vaikuttavat muun muassa datan rakenne, käytettävissä oleva tietomäärä ja tarve tulkita mallin päätöksiä. (Mucci, 2024)

Koneoppimiseen pohjautuvat ennustemallit ovat yleistyneet erityisesti liiketoimintaympäristöissä, joissa hyödynnetään suuria ja monimutkaisia tietomassoja. Esimerkiksi asiakaspoistuman ennustaminen, sairastumisriskien arviointi, luottoriskin hallinta ja tuotantolaitteiden vikaantumisen ennakointi ovat tyypillisiä sovelluksia (Brownlee, 2021). Kananen ja Puolitaival (Kananen, Puolitaival, 2019, ss. 127–128) painottavat, että tekoälyyn perustuvat ennustemallit tarjoavat yrityksille tehokkaita työkaluja muun muassa asiakaskäyttäytymisen analysointiin, varastohallinnan optimointiin sekä markkinoinnin kohdentamiseen.

Toimivan ennustemallin rakentaminen vaatii laadukasta ja monipuolista dataa, huolellista esikäsittelyä sekä mallin jatkuvaa testaamista ja päivittämistä. Ennustemallien onnistumista mitataan tyypillisesti tarkkuuden, herkkyuden tai muiden suorituskykymittareiden avulla (IBM, 2021) Vaikka ennustemallit tarjoavat merkittäviä hyötyjä päätöksenteon tueksi, niihin liittyy myös haasteita, kuten mallin ylisovittaminen (overfitting), vinoutunut data ja selitettävyyden puute. Kananen ja Puolitaival (2019) korostavat, että ennustemallien käyttö vaatii teknisen osaamisen lisäksi liiketoimintaympäristön ja eettisten näkökulmien ymmärtämistä.

2.1.2 Ennustemallin kehittämisen prosessi

Ennustemallin prosessissa on noudatettava tiettyjä vaiheita, jotta ennustetulokset olisivat mahdollisimman hyviä. Prosessin seuraaminen auttaa myös varmistamaan ennusteen uskottavuutta. (Sanders, 2015, s. 20)

Ennustemallin prosessin eri vaiheet:

1. Tunnista mitä ennustetaan

Ennustemallia rakentaessa etsitään ratkaisua johonkin kysymykseen. Ensimmäisessä vaiheessa tulee tietää mitä halutaan ennustaa ja millä tarkkuudella. Pintapuolisesti samankaltaiselta vaikuttavat kysymykset voivat kuulostaa samanlaisilta, vaikka niillä on eri merkitys lopputuloksen kannalta. Vaiheen päätöksillä on suuri vaikutus koko ennusteprosessin mallintamisessa. Ilmiön jakaantuminen lyhyelle tai pitkälle aikavälille vaikuttaa siihen mille aikavälille ennuste halutaan laatia. Data on poimittava mahdollisimman yksityiskohtaisella tasolla ja on valittava missä ja milloin uutta dataa kerätään ja mitä yksiköitä käytetään. Tavoitetta määriteltäessä johtamisen ja asiakastyön näkökulmien lisäksi myös operatiiviseen käyttöön liittyvät rajoitteet, kuten esimerkiksi tietosuojakysymykset. (Koivisto ym., 2023, s. 33)

2. Tunnista käyttäytymismallit ja määrittele tietoaineisto

Datan analysoinnissa pyritään hahmottamaan erilaiset toistuvat tapahtumat. Tärkeintä on tunnistaa, mitä kuvioita datassa esiintyy, jotta seuraavassa vaiheessa voidaan valita sopivin ennustemalli juuri tuolle datalle. Yleisimpiä kuvioita ovat esimerkiksi trendit, kausivaihtelut ja syklit. Lisäksi malleissa voi olla ennakoimatonta satunnaisvaihtelua ilman kaavaa, ja mitä enemmän sellaista on, sitä vaikeampaa ennustaminen on. (Sanders, 2015, s. 22) Aineiston määrittely on ennustemallin kehittämisen alkuvaiheen keskeinen tehtävä. Tekoälypohjaisten ennustemallien kehittämisessä keskeistä on ryhmien välisten erojen tunnistaminen päätepisteiden saavuttamisessa, joko ennalta määritettyjen ryhmien tai mallinnuksen itsensä avulla. Tämä edellyttää yhteistyötä asiantuntijoiden kanssa sekä ryhmien määrittelyssä että mallien tulosten tulkinnassa. Prosessin onnistumiseksi on tärkeää tunnistaa tavoitteen kannalta olennaiset tietoaineistot, arvioida niiden saatavuus ja laatu sekä varmistaa, että tarvittavat osaamisalueet, kuten data-analytiikka,

tietojärjestelmätuntemus ja käytännön palvelutyön tuntemus, ovat edustettuina. Tietoaineistojen hyödyntämistä rajoittavat usein datan saatavuus ja hyödynnettävyys aikataulun puitteissa. Aineistojen määrittelyssä on noudatettava muun muassa datan minimoinnin periaatetta ja varmistettava, että aineisto on riittävän laadukasta ja käyttökelpoista, koska puutteellinen tai väärin muotoiltu data voi heikentää mallin suorituskykyä. Prosessin kannalta on suositeltavaa rajata käytettävä data sellaisiin kokonaisuuksiin, joita voidaan tehokkaasti hyödyntää kehittämistyössä käytettävissä olevin resurssein. Ennustemalli on vain yhtä hyvä kuin sen syöttämä data. Datan tulee siis olla luotettavaa, tulevaisuutta ennustavaa, siivottua ja tarkistettua ennen sen käyttöä. Datan puhdistamisessa pyritään korvaamaan puuttuvat arvot sekä korjaamaan tai poistamaan huonolaatuinen data. (Koivisto ym., 2023, ss. 34–35; Sanders, 2015, ss. 20–21)

3. Aineiston hankinta

Aineiston hankinta tekoälypohjaisten mallien kehittämistä varten on monivaiheinen prosessi, jossa tulee huomioida erityisesti tietosuojalainsäädäntö, kuten GDPR ja toisiolaki. Ensimmäinen askel on määrittää, onko kyseessä organisaation omaan toiminnan kehittämiseen liittyvä hanke vai tieteellinen tutkimus ja käytetäänkö yhden vai useamman rekisterinpitäjän tietoja. Tämä vaikuttaa siihen, millaisia lupia ja järjestelyjä tarvitaan. Esimerkiksi kehittämistyö, jossa käytetään vain organisaation omia tietoja, voi edetä nopeammin, kun lupa voidaan myöntää sisäisesti. Prosessin sujumuuden kannalta on tärkeää suunnitella etukäteen, millaista dataa tarvitaan ja miten sen käyttö voidaan järjestää käytännössä. Eri toimijoidenvälinen data saattaa olla eri muodossa tai toimijoilla ei välttämättä ole kovinkaan tarkkaa tietoa toisten datasta. Kehitystyössä tarvitaan toimijoiden välisiä sopimuksia ja riskien hallintaan liittyviä dokumentteja. Tällaisia ovat esimerkiksi tietojenkäsittelysopimus (DPA), yhteisrekisterinpitäjien sopimus yhteishankkeissa sekä vaikutustenarviointi (DPIA), joka auttaa ennakoimaan mahdolliset tietosuojaan liittyvät riskit. Myös tietosuojaselosteet ja muut dokumentit kuuluvat prosessiin. Vaikka vaiheet voivat vaikuttaa byrokraattisilta, ne luovat pohjan sille, että dataa voidaan käyttää vastuullisesti ja luotettavasti mallien kehittämisessä. (Koivisto ym., 2023, ss. 37–38)

4. Ennustemallin valinta

Ennustamiseen on olemassa useita malleja ja valinta perustuu siihen, millaisia kuvioita datassa esiintyy, kuinka paljon ja millaista dataa on saatavilla, millä tarkkuudella ennustetta tarvitaan ja kuinka pitkälle tulevaisuuteen halutaan ennustaa. (Sanders, 2015, ss. 24–25)

5. Aineiston poimiminen

Aineiston poiminta alkaa vasta, kun tarvittavat luvat on saatu, ja se tarkoittaa käytännössä datan hakemista lähdejärjestelmistä tai erilaisista tietoaista. Tämän jälkeen alkaa esikäsittely, jossa aineisto käydään huolellisesti läpi. Tarkistetaan, että mukana on oikeat muuttujat, tiedot kattavat halutun ajanjakson ja ovat keskenään vertailukelpoisia. Tässä kohtaa aineiston laatu nousee keskiöön. Jos tiedot ovat puutteellisia, virheellisiä tai keskenään ristiriitaisia, niihin ei voi luottaa ja tällöin koko mallin rakentaminen vaarantuu. Siksi aineistoa tarkastellaan usein sekä sisällöllisesti että ajallisesti, ja sitä verrataan esimerkiksi tilastotietoihin, jotta mahdolliset puutteet tulisivat esiin. Jos tietoja puuttuu, niitä pyritään täydentämään tai käsittelemään järkevällä tavalla analyysivaiheessa. Lisäksi aineistoa voi olla tarpeen muokata tai yhdistellä eri lähteistä, etenkin jos mukana on useiden alueiden tietoja. Vaikka nämä vaiheet voivat viedä aikaa ja vaatia tarkkuutta, ne luovat pohjan koko mallin luotettavuudelle. (Koivisto ym., 2023, ss. 39–40)

6. Mallin toteutus ja testaaminen

Mallin toteutus ja testaus perustuvat pitkälti siihen, miten aineistoa käsitellään ja jaetaan eri vaiheisiin. Koneoppimisessa data jaetaan yleensä kahteen pääosaan, toista osaa käytetään mallin koulutukseen ja sen aikana tapahtuvaan arviointiin, ja toinen osa pidetään täysin erillään koulutusvaiheesta, jotta mallin toimivuutta voidaan lopuksi testata puolueettomasti. Joskus käytetään ristivalidointia, jossa mallia koulutetaan ja arvioidaan useammalla eri tavalla pilkokuilla aineistojoukoilla, tämä auttaa saamaan tarkemman kuvan mallin yleisestä suorituskyvystä. Jotta mallin arviointi olisi luotettavaa, on tärkeää pitää huoli siitä, että samaan henkilöön liittyvät havainnot eivät jakaudu eri aineistoryhmiin, muuten testitulokset voivat antaa harhaanjohtavan kuvan mallin toimivuudesta. Erityisesti, jos ennustettavia tapauksia on vähän, aineiston epätasapaino pitää huomioida esimerkiksi mallin painotuksia säätämällä. Myös monipuoliset arviointimittarit ovat tarpeen, jotta mallin suorituskyyä voidaan arvioida eri näkökulmista. Aikasarjadatan kohdalla on keskeistä, että mallin ennustepiste vastaa todellista käyttötarkoitusta. Tällöin mallin koulutusasetelman on oltava linjassa sen kanssa, miten mallia aiotaan käytännössä hyödyntää. Aineisto voidaan

esikäsitellä eri tavoin sen mukaan, millaista mallia rakennetaan. (Koivisto ym., 2023, ss. 41–42) Varsinainen ennusteen laatiminen on usein teknisesti yksinkertaista, sillä ennuste tuotetaan yleensä ohjelmiston avulla. Silti on tärkeää ymmärtää, mitä laskennan taustalla tapahtuu, millaista dataa on käytetty ja millä tavalla valittu malli toimii. Myös ohjelmiston tuottama lopputulos on merkityksellinen, sillä se toimii perustana seuraavalle vaiheelle ennusteprosessissa. (Sanders, 2015, s. 25)

7. Ennustemallin tarkkuuden mittaaminen

Kun ennuste on tehty ja todelliset tapahtumat toteutuvat, on erittäin tärkeää arvioida ennusteen osuvuutta mittaamalla ennustevirhe. Tätä tietoa tulisi käyttää ennustusprosessin kehittämiseen. On hyvä muistaa, että ennustaminen on jatkuva prosessi, joka elää ja kehittyy uuden tiedon ja datan myötä. (Sanders, 2015, ss. 20–25)

2.1.3 Datan rooli ennustemallin perustana

Data on raaka-ainetta, josta voidaan jalostaa merkityksellisempää tietoa, jota voidaan puolestaan välittää edelleen. Hyvinvointi- ja terveysdata on nousemassa reilun datatalouden keskeiseksi uusiutuvaksi raaka-aineeksi, jonka avulla voidaan kehittää entistä parempia digitaalisia palveluja sekä tehokkaampia, laadukkaampia ja yksilöllisempiä ratkaisuja ihmisten tarpeisiin. Arkistoitu data ei ole itsessään hyödyllistä, ellei sitä voida jalostaa päätöksenteon, ennustamisen ja analytiikan tueksi. Yhteiskäyttö eri toimijoiden kesken edellyttää yhteensopivia formaatteja ja järjestelmiä, luotettavuuden ja eheyden varmistamista, sekä selkeitä sääntelykehikkoja ja tietosuojan noudattamista. (Benke & Benke, 2018, Osat 6–8; Hendolin & Hämäläinen, 2022)

Big data -analytiikka sosiaali- ja terveydenhuollossa perustuu neljään keskeiseen ominaisuuteen: määrä (volume), nopeus (velocity), monimuotoisuus (variety) ja luotettavuus (veracity). Nämä ominaisuudet kuvaavat sekä datan käsittelyn haasteita että mahdollisuuksia. Dataa syntyy jatkuvasti valtavia määriä eri toimijoilla, ja haasteena on datan tallentaminen, hallinta ja käsittely nykyaikaisia alustoja käyttäen. Datan järkevä hyödyntäminen vaatii jo alkujaan tehokkaita menetelmiä sekä laskentatehoa ja hajautettua prosessointia. Datan täytyy olla käytettävissä ja analysoitavissa lähes reaaliajassa, jotta sillä voi olla vaikutusta asiakkaiden ohjaukseen ja päätöksentekoon. Datan jatkuva virta vaikeuttaa sen esikäsitelyä, puhdistusta ja laadunvarmistusta ennen analyysiä. Dataa voi

myös olla monessa muodossa, helposti luettavissa ja käsiteltävissä olevaa dataa, sähköisiin järjestelmiin vietyä, mutta osittain vapaamuotoista tai tulkinnanvaraista tekstiä, sekä vapaamuotoisia muistiinpanoja ja tietoja, kuten keskustelut, kuvat yms. Useimmat analytiikkajärjestelmät on suunniteltu rakenteelliselle datalle, mutta suuri osa arvokkaasta tiedosta on edelleen väärässä muodossa käytettäväksi. Datan yhdistäminen eri lähteistä vaatii myös usein muunnoksia, esimerkiksi eri organisaatiot käyttävät eri käsitteitä ja luokituksia. Eettisiä ja lainsäädännöllisiä datankäytön rajoituksia käsitellään luvussa 4. (Benke & Benke, 2018, Osat 6–8; Raghupathi & Raghupathi, 2014, ss. 3–4)

3 TYM-yhteistoimintamalli

Työllistymistä edistävä monialaisen tuen yhteistoimintamalli (TYM-yhteistoimintamalli) on lakisääteinen Kelan, työvoimaviranomaisen ja hyvinvointialueen yhteinen toimintamalli. Työvoimauudistus 1.1.2025 siirsi julkisten työvoimapalveluiden järjestämisvastuun TE-toimistoilta kunnille. Toimijat vastaavat yhdessä työttömän palvelutarpeen arvioinnista ja hänelle soveltuvien palvelujen suunnittelusta. Mallin tarkoitus on hyödyttää työtöntä asiakasta, jonka työllistyminen edellyttää viranomaisten palvelujen yhteensovittamista ja työllistymisen tukitoimia. Työvoimaviranomainen vastaa yhdessä hyvinvointialueen sekä tarvittaessa Kelan kanssa asiakkaan monialaisten palvelutarpeiden arvioinnista, tarkoituksenmukaisen palvelukokonaisuuden rakentamisesta ja työllistymisprosessin ohjauksesta ja seurannasta. Moniallinen työllistymissuunnitelma laaditaan työvoimaviranomaisen, hyvinvointialueen ja asiakkaan yhteistyönä, ja siihen kirjataan kaikki sovitut toimenpiteet. Kela osallistuu suunnitelman valmisteluun aina, kun asiakkaalle on tarpeen järjestää sen tuottamaa kuntoutusta. Monialaisen yhteispalvelun tarkoituksena on edistää työttömien työllistymistä tarjoamalla palvelutarpeen mukaisia julkisia työvoimapalveluja sekä sosiaali-, terveys- ja kuntoutuspalveluja. (Kela, 2025; Liski-Wallentowitz, 2024, ss. 5–6)

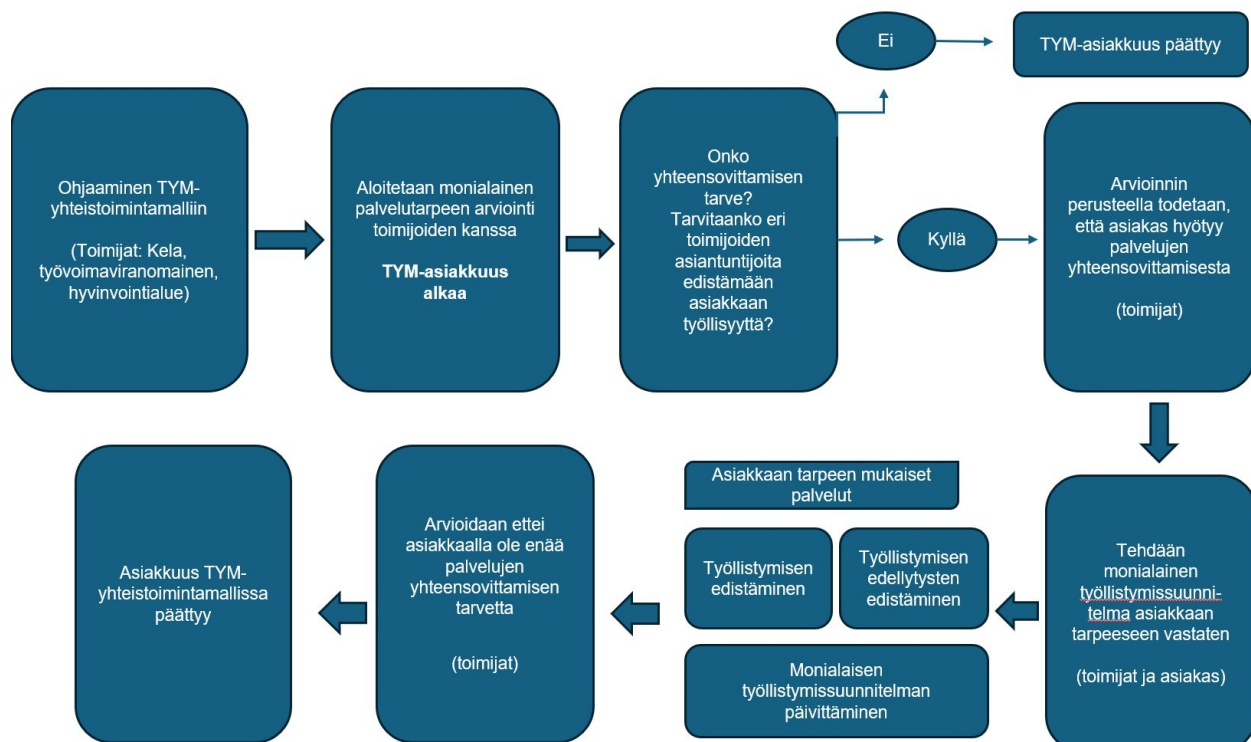
3.1 TYM-mallin toimijat ja palveluprosessi

Työvoimapalvelut siirtyivät kuntien järjestämisvastuulle 1.1.2025. Osana TE-palvelut 2024-uudistuksen toimeenpanoa yhdyspintojen yhteistyön rakentamiseksi on annettu ”Kansalliset suositukset yhdyspintojen rakentamiseen alueille”. Yhdyspinnalla viitataan eri organisaatioiden väliseen toiminnalliseen rajapintaan. Tässä työssä sillä tarkoitetaan

toiminta-alueella, jolla työvoimaviranomainen, Kela ja hyvinvointialue toimivat yhdessä työllistymistä edistävien palvelujen parissa. Tällä alueella viranomaisten palvelut ja asiakkaan palvelupolut sovitetaan yhteen, jotta tuki muodostaa mahdollisimman yhtenäisen kokonaisuuden. Yhdyspinnan asiakkaalla taas tarkoitetaan työkäistä henkilöä, joka työllistymisensä edistämiseksi tarvitsee useamman viranomaisen palveluja parantaakseen osallisuuttaan työelämässä. (Kuntaliitto, 2024; Valtioneuvosto, 2024) Koska asiakkaiden tarpeet ylittävät usein organisaatorajat, ei yksittäinen taho voi vastata moninaiisiin palvelutarpeisiin. Toimivat yhdyspinnat tulee rakentaa jo uudistuksen valmisteluvaiheessa, jolloin viranomaisyhteistyön merkitys korostuu. (Valtioneuvosto, 2024, ss. 6–7)

Asiakasohjauksessa on olennaista tunnistaa monialaisen tuen tarve ja sen edellyttämä palveluprosessi. Toimivat palveluprosessit mahdollistavat asiakkaan tarvitsemat palvelut oikea-aikaisesti. (Kuntaliitto, 2024) TYM-yhteistoimintamallin eri vaiheet (Kuva 1) kuvaavat työllistymistä edistävän monialaisen tuen perusprosessia, kuten se on määritelty TYM-laissa. (KEHA-keskus, 2024, ss. 6–7)

Kuva 1. TYM-prosessi mukaillen (KEHA-keskus, 2024)



Kun TYM-asiakkuus alkaa, avataan ja aktivoidaan asiakkuus TYM-yhteistoimintamallin TYPPI-asiakastietojärjestelmässä. Järjestelmää hyödynnetään tiedon jakamisessa ja suunnitelmien kirjaamisessa. Kun TYM-asiakkuus päättyy, jatketaan asiakkuutta tarvittaessa muissa palveluissa. (KEHA-keskus, 2024, ss. 6–7)

3.2 TYM-laki

Laki työllistymisen monialaisesta edistämisestä (L 381/2023) tuli voimaan 1.1.2025 alkaen. Lain tavoitteena on edistää vaikeassa työmarkkina-asemassa olevien henkilöiden työllistymistä tarjoamalla monialaista tukea ja säädellä viranomaisten välistä yhteistyötä tilanteissa, joissa työtön henkilö tarvitsee useiden eri palvelujen yhdistelmää päästäkseen työelämään tai koulutukseen. (Laki työllistymisen monialaisesta edistämisestä 381/2023 § 3)

Monialaisen tuen yhteistoimintamallissa työvoimaviranomainen, hyvinvointialue ja Kansaneläkelaitos arvioivat yhdessä työttömän palvelutarpeen, suunnittelevat ja yhteen sovittavat tarkoituksenmukaiset palvelut sekä vastaavat prosessin etenemisestä ja seurannasta. Mallissa yhdistetään työvoimapalvelut sekä sosiaali-, terveys- ja kuntoutuspalvelut, joiden järjestämisestä säädetään muussa lainsäädännössä. (Laki työllistymisen monialaisesta edistämisestä 381/2023 § 2).

Lain mukaan Kansaneläkelaitos, hyvinvointialue sekä työvoimaviranomainen voivat käsitellä työtöntä henkilöä koskevia tietoja moninaisen tuen yhteistoimintamalliin liittyvien tehtävien hoitamiseksi, sekä valvontaan, kehittämiseen, seurantaan, arviointiin, tilastointiin, ennakkointiin ja ohjaukseen liittyen. Henkilöä koskevat tiedot ovat nimi, henkilötunnus ja yhteystiedot, yhteistoimintamallin asiakkuuteen ja asiointiin liittyvät tiedot, koulutusta ja työhistoriaa koskevat tiedot, sosiaalista tilannetta sekä terveydentilaa koskevat tiedot sekä palvelutarvetta ja työllistymistä koskevat tiedot. Tietojenkäytön edellytys on, että ne ovat välttämättömiä palvelun tarjoamiseksi. (Laki työllistymisen monialaisesta edistämisestä 381/2023 § 11)

Kehittämisen- ja hallintokeskus KEHA ylläpitää TYM-yhteistoimintamallin asiakastietoaineistoa sekä asiakastietojärjestelmää. Työvoimaviranomaisen, hyvinvointialueen ja Kansaneläkelaitoksen on talletettava 11 §:ssä tarkoitetut tiedot monialaisen tuen yhteistoimintamallin asiakastietojärjestelmään osaksi

asiakastietoaineistoa. Tietoja saa tallentaa monialaisen tuen yhteistoimintamallin asiakkuuden alkamisen jälkeen, asiakkuus alkaa, kun palvelutarpeen arviointi aloitetaan. Tiedot on poistettava, kun asiakkuuden päättymisestä on kulunut neljä vuotta. (Laki työllistymisen monialaisesta edistämisestä 381/2023 § 4, § 12)

Kansaneläkelaitos, hyvinvointialue ja työvoimaviranomainen ovat tietojen yhteisrekisterinpitäjiä ja saavat tallettaa 11§:n 2 momentissa tarkoitettuja tietoja siltä osin, kun ne ovat tarpeellisia palvelun tarjoamiseksi työttömälle. Viranomaiset voivat käyttää laissa säädettyjen tehtäviensä hoitamiseksi työtöntä koskevia tietoja, jotka ovat välttämättömiä työllistymistä edistävän monialaisen tuen yhteistoimintamallin järjestämiseksi. (Laki työllistymisen monialaisesta edistämisestä 381/2023 § 13, § 14, § 15) Laki edellyttää, että viranomaiset sopivat keskenään pysyvästä yhteistoimintamallista, jonka avulla varmistetaan sujuva yhteistyö asiakkaan tueksi. Viranomaisten yhteistyö ei siis perustu vapaaehtoisuuteen, vaan se on säädetty lakisääteiseksi velvollisuudeksi. (*Laki työllistymisen monialaisesta edistämisestä* | 381/2023 | *Lainsäädäntö* | *Finlex*, 2025)

4 Digitalisaatio julkishallinnossa

Pääministeri Petteri Orpon hallitusohjelman mukaan hallitus toteuttaa perusturvan uudistuksen, jonka tavoitteena on parantaa työnteon kannattavuutta, sujuvoittaa sosiaaliturvaa sekä yksinkertaistaa etuuksia. Uudistuksen yhteydessä perustettiin asiantuntijaryhmä kartoittamaan digitalisaation mahdollisuuksia. (Sosiaali- ja terveystieteiden ministeriö, 2025) Digitalisaatiolla on keskeinen rooli sosiaaliturvauudistuksessa, sillä sen avulla pyritään edistämään yhdenvertaisuutta ja tehostamaan julkisen hallinnon toimintaa. Digitaaliset ratkaisut voivat parantaa viranomaisten välistä yhteistyötä ja tukea tietoperustaista johtamista. Samalla on tunnistettu myös haasteita, kuten digitaalinen eriarvoisuus, kaikilla ei ole mahdollisuutta asioida digitaalisesti, joten vaihtoehtoiset asiointitavat on säilytettävä ja yksilölliset tarpeet huomioitava. Sosiaaliturvajärjestelmän ja digitalisaation haasteena on järjestelmän monimuotoisuus. Toimeenpanoon osallistuu suuri joukko toimijoita, joilla on erilaisia käytäntöjä, ja samalla järjestelmän on pystyttävä vastaamaan ihmisten vaihteleviin elämäntilanteisiin. Erityisen haastavaa on yhdistää etuuksien yhdenvertainen kohtelu palvelujen yksilölliseen ja asiakaslähtöiseen toteutukseen. (Spisak, 2024, ss. 10–11)

Suomessa Sosiaali- ja terveysministeriö on julkaissut digitalisoimisen tiekartan 2023–2027. Tiedonpuutteen vuoksi digitalisaation täyttää potentiaalia ei ole vielä tähän mennessä pystytty hyödyntämään työkyky- ja kuntoutuspalvelujen hankkeissa. Tiekartan pyrkimyksenä on korjata tämä ongelma kuvaamalla digitalisaation mahdollisuuksia. Kehittämällä tiedonkulkua toimijoiden välillä, voidaan tehokkaammin tunnistaa ongelmia, edistää palveluketjujen toteutumista ja mahdollistaa etuusjärjestelmien tietoon perustuvaa johtamista. (Sosiaali- ja terveysministeriö Helsinki, 2023, s. 7) Julkisissa palveluissa syntyy runsaasti dataa, jonka tehokas hyödyntäminen esimerkiksi palvelujen suunnittelussa, johtamisessa ja tutkimuksessa on kuitenkin haastavaa. Yksi keskeinen ongelma on datan hajanaisuus. Viranomaisilla on toisistaan poikkeavia tietorakenteita, jolloin datan yhdisteltävyys ja vertailukelpoisuus on vaikeaa. Toimijat keräävät dataa samoista asiakkaista omiin tietojärjestelmiinsä omassa muodossaan. Esteitä liittyy lisäksi datan laatuun, ajantasaisuuteen, kattavuuteen, luotettavuuteen ja ylipäättään tarpeellisen datan puuttumiseen. (Salovaara ym., 2021, ss. 379–380) Laadukas data on edellytys toimiville tekoälymalleille, joten sen ylläpitoon on panostettava jatkuvasti. Ennustemallia kehittäessä määritellään, mitä dataa eri tietojärjestelmistä tarvitaan, joten organisaatiossa on oltava kyky tunnistaa tietojärjestelmissä tuotetun tiedon ja siitä saatavan datan luotettavuutta. Suomessa yhtenä haasteena on myös se, että tietoja tallennetaan eri järjestelmiin, jotka eivät ole vuorovaikutuksessa keskenään eikä muiden dataa tunneta. (Koivisto & Tiirinki, 2020, s. 16)

Tämän lisäksi tietosuojalainsäädännön monimutkaisuus voi estää tiedon tarkoituksenmukaisen hyödyntämisen. Uudistunut lainsäädäntö saatetaan myös kokea vaikeaksi. Tiedon kerääminen kuormittaa ammattilaisia, ja ilman selkeitä hyödyntämistarkoituksia tieto ei tue päätöksentekoa riittävästi. Kehittämisen lähtökohtana tulisi olla tiedon konkreettisten käyttötarkoitusten tunnistaminen. Tiedon tuottamiseen ja jalostamiseen käytettävät resurssit on kohdennettava sinne, missä vaikutukset ovat merkittävimmät. Erityisesti työkyky- ja kuntoutuspalvelujen sekä niihin liittyvien etuusprosessien osalta tietoa voitaisiin hyödyntää vaikuttavuuden seurannassa, palveluiden yhteensovittamisessa ja monialaisen yhteistyön tehostamisessa. Tietojohdamisessa tulee huomioida laaja-alaisesti lainsäädännön kehittämistarpeet, tietojärjestelmien kehittämistarpeet sekä tietoa tuottaville ammattilaiselle aiheutuva lisätyö. (Sosiaali- ja terveysministeriö Helsinki, 2023, ss. 9–11)

4.1 Dataekosysteemit ja data-avaruus

Dataekosysteemi on useiden toimijoiden välinen kumppanuus, jossa jaetaan sekä hallinnoidaan dataa uudenlaisen arvon tuottamiseksi. Datan jakaminen perustuu vastavuoroisuuteen, minkä vuoksi eri lähteistä tuleva data rikastuttaa tiedon monimuotoisuutta ja lisää osallisuutta myös yksilöille, kuluttajille ja kansalaisille. Tyypillisessä dataekosysteemissä useat organisaatiot, tiedon tuottajat, kokoajat ja käyttäjäorganisaatiot toimivat yhdessä edistääkseen datan jakamista ja tuottaakseen lisäarvoa. (Capgemini, 2021)

Suomessa Seinäjoen kaupungin työllisyyspalvelut ja Vastuu Group rakentavat valtakunnallista työllisyyspalveluiden dataverkosta. Työnantajat ja työllistymistä edistävät palvelut ovat työnhakijan saavutettavissa tämän verkoston kautta. Dataverkoston tarkoitus on yhdistää yksityinen ja julkinen sektori, jolloin yksi yhteinen dataverkosto hyödyttää kaikkia työllisyysalueita ja niiden työnhakijoita. Toteutuessaan dataverkosto auttaa järjestämään ja selkeyttämään paikallisen työllisyyden kehittämistä ja johtamista. Se auttaa asiakasta suunnistamaan palveluviidakossa ja suuntautumaan oikeiden palvelujen piiriin. Toimiva dataverkosto selkeyttää palveluntarjoajan kokonaiskuvaa ja vahvistaa yhteistyömahdollisuuksia yli rajojen. Datanvälityspalvelussa käyttäjällä on mahdollisuus turvallisesti jakaa tietojaan ja antaa suostumuksiaan alueen toimijoille. Tarvitaan työnhakijan tiedot ja suostumus niiden käyttöön, jotta toimijoiden on mahdollista tarjota työnhakijalle oikea-aikaisesti työtä, palveluita tai koulutusta. (Vastuu Group, 2023, 2025)

Tietoon perustuva suostumus on tärkeä ja paljon keskustelua herättänyt aihe sosiaali- ja terveysdata hyödyntämisessä. Tietoon perustuva suostumus on perinteisesti nähty tärkeänä tapana suojella yksilön oikeuksia, erityisesti silloin kun kerätään henkilötietoja. Ajatuksena on, että suostumuksen antaja ymmärtää, mihin hän suostuu, esimerkiksi millaisia vaihtoehtoja hänellä on ja mitä seurauksia niihin liittyy. Suostumuksen on myös perustuttava vapaaehtoisuuteen. Lisäksi edellytetään, että henkilö on kykenevä tekemään harkitun päätöksen. Massadatan ja kehittyneen analytiikan käyttö tuo uusia haasteita perinteiselle tietoon perustuvalla suostumuksella. Suostumus on yleensä annettu tiettyä tarkoitusta tai tutkimusta varten, mutta massadatan ideana on yhdistää tietoa eri lähteistä ja hyödyntää sitä myös ennakoimattomilla tavoilla. Mahdollisesti jopa niin, että alkuperäinen datan keruun käyttötarkoitus on eri kuin lopullinen käyttökohde. Tällöin yksilön on vaikea ymmärtää, mihin kaikkeen hän oikeastaan suostuu. Joskus ei edes datan

kerääjä tiedä vielä, mihin kaikkeen tietoa tullaan käyttämään. Siksi puhutaan siitä, että massadata haastaa koko suostumuksen idean, miten voi suostua johonkin, jota ei vielä ole määritelty? (Spisak, 2024, ss. 25–26)

Jotta data lähtisi sujuvasti liikkeelle, tulee organisaatioissa olla omat sisäiset datakyvykkyydet kohdillaan. Pyramidissa (Kuva 2) esitellään datataloudessa tarvittavien kyvykkyyksien eri alueita. Pyramidin ensimmäinen porraskuvaa sitä, että jos organisaatiot eivät tiedä itse mitä dataa heillä on tai missä järjestelmissä data on, mitä datalla voisi tehdä ja mikä sen käyttöarvo olisi, on vaikea lähteä kehittämään datapohjaista yhteistyötä muiden toimijoiden kanssa. Organisaation sisäisten datakyvykkyyksien kehittämistarpeet kumpuavat yleensä halusta tehostaa prosesseja, automatisoida toistuvia tehtäviä sekä hyödyntää talous- ja liiketoimintatietoa nykyistä tehokkaammin. Tässä vaiheessa ei vielä ole tarvetta data-avaruuksille, sillä toiminta ja tiedonhallinta tapahtuvat organisaation sisällä. Toinen porraskuvaa sitä, että ymmärretään miten omat järjestelmät auttavat haluamassamme prosessissa tai osaamisen ennakoinnissa, jos kuitenkin ei ole kytköksiä muihin toimijoihin. On ymmärrettävä miten dataa verkostoissa jakamalla ja muiden toimijoiden dataa hyödyntämällä voidaan parantaa asiakaskokemusta ja palveluita. Tällä tasolla on tärkeää osata tehdä verkostomaista yhteistyötä dataekosysteemissä sekä rakentaa luottamusta toimijoiden välille. Kolmas porraskuvaa siihen, että pystytään tekemään yhteistyötä ekosysteemeissä ja vaihtamaan dataa eri organisaatioiden välillä hallitusti. Suomessa on satoja eri dataekosysteemejä, ja vaikka kaikki eivät liity työllistämisen edistämiseen, datatalouden kyvykkyyksien periaate on kuitenkin kaikissa dataekosysteemeissä sama. (Vastuu Group, 2025)

Kuva 2. Datataloudessa tarvittavia kyvykkyyksiä mukaillen (Vastuu Group, 2025)



Suurin osa olemassa olevista dataekosysteemeistä on pieniä muutaman toimijan välisiä, ja toimijat ovat keskenään tehneet juridisen ja teknisen työn siinä mitä dataa voidaan vaihtaa toimijoiden välillä. Data-avaruudet ovat sen sijaan sellaisia, että niihin olisi uusien toimijoiden helppo liittyä mukaan, tuoda dataa sekä käyttää olemassa olevaa dataa. Sitran julkaisun (Poikola ym., 2024) mukaan ”data-avaruus on työkalu ekosysteemeille, jonka avulla ekosysteemit toimijat voivat siirtää dataa luotettavasti ja turvallisesti”. Sopimusjuridista tien silloittamista tarvitaan, data-avaruus on työkalu dataekosysteemeille jouhevaan datan jakamiseen. Datan oikeuksien haltijat päättävät datan käytöstä sekä siihen liittyvistä ehdoista, päätöksentekijöitä voivat olla henkilöt omasta henkilödatasta tai yritykset ja organisaatiot. Data-avaruuteen osallistuminen ei pakota datan avaamiseen, vaan kontrolli säilyy aina sillä, kenen datasta on kysymys. Data-avaruudessa on valmiita sopimusrakenteita siihen, miten dataa jaetaan. Se on hajautettu järjestelmä, ei ainoastaan kahden tai muutaman toimijan räätälöity ratkaisu. (Vastuu Group, 2025)

Euroopan komissio on rahoittanut 14 eri toimialakohtaisen data-avaruuden kehittämistä. Data-avaruudet tarjoavat mahdollisuuden vastata sellaisiin ongelmiin, joita mikään yksittäinen toimija ei pystyisi itsenäisesti ratkaisemaan. (”Suomalaisten data-avaruuksien tilannekuva”, 2024)

4.2 SOTE-tekoälyn ekosysteemi

Suomessa SOTE-tekoälyn ekosysteemiverkostoa johtaa sosiaali- ja terveysministeriö ja jäsenorganisaatioita on jo yli 250. Ekosysteemi on sosiaali- ja terveysalan toimijoista, kuten viranomaisista, yrityksistä ja tutkijoista, muodostuva vapaamuotoinen verkosto, joka rahoittaa, mahdollistaa ja selvittää tekoälyn käyttömahdollisuuksia alalla.

Ekosysteemiverkoston tarkoituksena on koota toimijoita yhteen kehitysprojektien toteuttamiseksi ja myöntää avustuksia. Samalla on mahdollisuus kerätä ja jakaa tietoa ja käyttötapauksia, sekä edelläkävijöiden kokemuksia. Tarkoituksena on kansallisten selvitysten lisäksi tunnistaa lainsäädännöllisiä muutostarpeita ja tehdä niiden osalta mahdollistavia ehdotuksia. Ekosysteemin tavoitteena on varmistaa asiakkaille laadukkaat palvelut, tukea ammattilaisia heidän työssään sekä vapauttaa aikaa keskeisiin tehtäviin, kuten henkilökohtaiseen asiakaskohtaamiseen. Lisäksi tekoälyteknologiat parantaisivat tuottavuutta kuitenkin turvallisuudesta ja yksityisyydensuojasta huolehtien. (DigiFinland, 2024a; Innokylä, 2025)

SOTE-tekoälyn ekosysteemin tekoälyvisio 2035 kuvastaa kuinka tekoäly on sulautunut ihmisen arkeen, palvelutuotantoon sekä johtamiseen. Automaatio, ennaltaehkäisy ja omatoimisuus korostuvat erityisesti terveydenhuollon palveluissa. Tekoälyn tarkoituksena olisi helpottaa asiointia ja lisätä asiakkaan omaa roolia hyvinvoinnin edistäjänä, mahdollistaa taloudellisesti kestävää toimintaa sekä lisätä toimijoiden yhteistyötä ja nopeuttaa sote-kehittämistä. Tekoälyvision toteuttamiseksi tarvitaan eri toimijoiden yhteistyötä sote-tietorakenteiden yhtenäistämiseksi, toimenpiteitä datan yhteiskäytön helpottamiseksi sekä automatisoitavien palvelupolkujen tunnistamiseksi. Kilpailutus tulee uudistaa tukemaan tekoälyinnovaatioiden nopeaa käyttöönottoa ja on luotava mittarit yhdenvertaisuuden ja eettisyyden varmistamiseksi. (Kallio-Könnö, 2025, s. 7,20,38)

4.3 Eettinen tekoälyn käyttö

Tekoäly on kehittynyt yhä paremmaksi auttamaan päätöksenteossa ja ohjaamaan toiminnoissa ja sen käyttö on sitä myötä yleistynyt. Tekoälyä käytetään tukemaan petosten havaitsemista, luottoriskin arviointia, terveydenhuollon diagnostiikkaa, rekrytointia ja paljon muuta. Näillä päätöksillä on suuri vaikutus yksilöiden elämään, jonka vuoksi tekoälyn käyttö on erittäin merkityksellinen osa arkea. Tekoälyjärjestelmät voivat tuottaa syrjiviä päätöksiä

tai analyyseja, koska niiden koulutusdata heijastaa menneitä olosuhteita ja mahdollisesti sisältää rakenteellista epätasa-arvoa. Esimerkiksi rekrytoinnissa, petosten tunnistamisessa ja sairausvakuutuksissa on ollut tapauksia, joissa tekoälyn käyttö on johtanut syrjintään etnisen taustan vuoksi. Kaikissa näissä tapauksissa tekoäly on ollut osana sosioteknistä järjestelmää, jossa uudet teknologiat ovat vuorovaikutuksessa sosiaalisten toimijoiden kanssa. Eettiset haasteet nousevat esiin teknologian tasolla mutta myös uusien sosioteknisten järjestelmien osalta. (DigiFinland, 2024b, s. 9; Smuha, 2025, ss. 59–60)

”Sosioteknisiä systeemejä ovat esimerkiksi modernit organisaatiot, joiden ydintehtävien hoitamiseksi tarvitaan ihmistoimijoiden ja monien teknologisten järjestelmien sujuvaa yhteistyötä ja joiden vakaa toiminta on olennaista turvallisuuden takaamiseksi.” (Hallamaa & Kalliokoski, 2023)

Esiin nousee monia moraalisia kysymyksiä, joita on tarkasteltava sosiaalisessa, psykologisessa, taloudellisessa ja yhteiskunnallisessa kontekstissa. On olennaista ymmärtää huolellisesti sitä aluetta, johon eettistä teoriaa sovelletaan. Meidän tulee tietää, millaisia ominaisuuksia tekoälytoimijoilla on ja miten ne eroavat ihmisestä. Monet kysymyksistä ovat jo ennestään tunnettuja, ne vaan nousevat esiin, kun tekoäly integroidaan korkean vaikutuksen prosesseihin, jotka ovat olleet jo olemassa ennen tekoälyä. Esimerkiksi syrjintää on tutkittu laajasti, samoin kuin siihen liittyviä oikeudenmukaisuuden ja tasapuolisuuden käsitteitä. (Smuha, 2025, s. 60)

Tekoäly eroaa muista teknologioista ainakin kahdella tavalla. Tekoälyllä voi ensinnäkin olla suurempi toimijuuden aste kuin muilla teknologioilla, ne voivat periaatteessa tehdä päätöksiä itsenäisesti ympäristöön reagoiden. Vähintään voidaan sanoa, että ne käynnistävät tapahtumaketjuja, joita ei olisi tapahtunut ilman niiden aloitetta. Syväoppimisjärjestelmien kohdalla emme kuitenkaan aina tiedä, miksi järjestelmät tuottavat juuri tietyn tuloksen. Tietojenkäsittelytieteilijät pystyvät selittämään järjestelmän toimintaa ja yleisiä piirteitä, kuitenkin yksittäisille tuloksille ei aina ole selitystä. Nämä ominaisuudet tekevät tekoälyjärjestelmien kehittämisestä, käyttöönotosta sekä vastuullisesta käytöstä haastavaa. Vaikka tekoälyllä on enemmän toimijuutta kuin muilla teknologioilla, ei sillä kuitenkaan ole moraalista toimijuutta, eli se ei voi kantaa vastuuta tuottamiensa tulosten seuraamuksista. Tekoälyjärjestelmät ovat usein niin monimutkaisia ja läpinäkymättömiä, että jopa niiden kehittäjien on vaikea ennustaa, mitä ne tekevät ja miksi. Tämän seurauksena tutut eettiset ongelmat, jotka syntyvät vastuuttomasta tai väärin kohdistetusta

toiminnasta, toistuvat ja pahenevat tekoälyjärjestelmien tuoman nopeuden, laajuuden ja läpinäkymättömyyden vuoksi. Tämän takia on vaikeaa käyttää tekoälyä vastuullisesti ja varmistaa, että se toimii eettisesti oikein. Siksi on tärkeää löytää keinoja, joilla eettiset arvot saadaan osaksi tekoälyn toimintaa, vaikka se onkin haastavaa. (Smuha, 2025, ss. 61–63)

Tekoälyn toiminnan onnistumista voidaan tarkastella kahdesta eri näkökulmasta, tehtävä suoritetaan suunnitellusti ja tavoiteltu vaikutus toteutuu. Kun tarkastellaan sosioteknisten järjestelmien toimintaa, on olennaista tehdä ero sen välillä, toimiiko järjestelmä sinänsä hyvin, ja sen välillä, saavutetaanko sillä asetetut tavoitteet. Eli vaikka itse teko suoritetaan onnistuneesti, ei välttämättä saavuteta tavoitetta. Toiminnan onnistuminen ei aina tarkoita, että se olisi hyvä tai oikeudenmukainen muiden kannalta. Yhteistyön voidaan sanoa onnistuneen, kun suunnitellut teot tehdään oikein ja tavoite saavutetaan. Mutta lisäksi olisi hyvä, että toiminta tuottaa myös sellaisia asioita, joita toimijat pitävät arvokkaina ja hyvinä. Tilanne muuttuu silloin, jos toiminta epäonnistuu ja se aiheuttaa haittaa tai epäedullisia seurauksia muille osapuolille. Tällöin syntyy tarve selvittää syitä ja syyllisiä ja löytää taho, jota voidaan vaatia vastuuseen. Epäonnistumisen horjuttavat luottoa toimijoihin ja heidän moraaliseen ja juridiseen vastuuseen. (Hallamaa & Kalliokoski, 2023, ss. 141–142)

Vastuullisuutta ja vastuuta on pidetty perinteisesti ihmisen ominaisuutena, koska heitä voi moittia, edellyttämään pitämään lupauksensa ja vaatia vastuuseen. Tekoölyyn nämä vastuun vaatimukset eivät kuitenkaan sovi, koska sillä ei ole tietoisuutta, omia tavoitteita tai moraalitunteita, kuten ylpeyttä tai häpeää. Tekoälyä arvioitaessa parempi vertailukohta kuin yksilö on kollektiivinen toimija ja sen vastuullisuuden tarkastelu. Tekoälysystemejä ja ihmisryhmiä, kuten yrityksiä tai organisaatioita, yhdistää kysymys siitä, voidaanko niitä pitää toimijoina ja pitääkö niitä vastuullisina. Ryhmätoimijoita ovat esimerkiksi yritykset ja viranomaiset, joiden nimissä useat ihmiset voivat toimia ilman yksilökohtaista vastuuta. Ne voivat yhdessä muodostaa näkemyksiä ja tavoitteita ryhmänä. Näilläkään ei ole olemassa niin sanottua ryhmämieltä, vaan toiminta perustuu aikomuksiin, päätöksiin ja aikaansaannoksiin. Vastaavalla tavalla tekoälysystemi päättää ja aikaansaa sille opetettujen sääntöjen mukaisesti. (Kalliokoski, 2023) Ryhmätoimijat ja tekoäly voivat saada aikaan seurauksia, joista kukaan yksittäinen toimija ei tunne vastuuta, tätä kutsutaan vastuuaukoksi. Tekoäly voi vaikuttaa tarkoitukselliselta toimijalta, mutta siltä puuttuu tietoisuus ja moraalit. Vastuuaukot herättävät oikeudenmukaisuushuolia, ja siksi filosofit ovat alkaneet kehittää uusia käsitteellisiä tapoja määritellä vastuullisuutta myös tekoälyn ja kollektiivien kohdalla. (Hallamaa & Kalliokoski, 2023, ss. 148–150)

OECD eli Taloudellisen yhteistyön ja kehityksen järjestö on laatinut tekoälyn eettiset periaatteet, jotka ovat maailman ensimmäisiä hallitustenvälisiä suosituksia tekoälyn vastuullisesta käytöstä. Näitä periaatteita noudattavat monet maat ja ne ovat vaikuttaneet myös EU:n tekoälyasetuksen sisältöön. Läpinäkyvyys ja selitettävyys ovat keskeisiä periaatteita OECD:n luotettavan tekoälyn periaatteissa, jolloin automaattiseen päätöksentekoon tulisi sisältyä ihmisen osallistumista siinä määrin kuin tekoälyn tuottamien lopputulosten vaikutus edellyttää. tekoälytoimijoiden tulisi ottaa käyttöön mekanismeja ja turvatoimia, kuten mahdollisuus ihmisen tekemään lopulliseen ratkaisuun. (OECD, 2025, s. 19)

4.4 Tekoälyn käytön keskeinen lainsäädäntö

Tässä luvussa käsitellään tekoälyn käytön keskeisintä lainsäädäntöä ennustemallien ja datanjakamisen näkökulmasta. Lainsäädäntöä on tarkasteltu sen osalta, mitä erityisesti on huomioitava ennustemallin kehittämisessä, listaus ei kuitenkaan ole tyhjentävä. Luvussa 3.4. on käsitelty tarkemmin TYM-lakia ja sen määrittelemiä oikeuksia ja velvoitteita.

Euroopan parlamentin ja neuvoston asetus (EU General Data Protection Regulation 2016/679) luonnollisten henkilöiden suojasta henkilötietojen käsittelyssä ja näiden tietojen vapaasta liikkuvuudesta, eli yleinen tietosuoja-asetus (GDPR), tuli voimaan 25.5.2018. Asetusta sovelletaan henkilötietojen käsittelyyn, ja se määrittää yleiset periaatteet ja edellytykset henkilötietojen käsittelylle. Henkilötietojen käsittely kattaa lähes kaikki toimenpiteet, joita tehdään henkilötiedoille, joko automaattisesti järjestelmien avulla, tai käsin työntekijän toimesta. Tällaisia toimenpiteitä ovat esimerkiksi tietojen kerääminen, tallentaminen, säilyttäminen, muokkaaminen, käyttäminen, siirtäminen, luovuttaminen tai poistaminen. Henkilötietojen käsittelyssä on aina noudatettava tietosuoja-asetuksessa määritellyjä periaatteita, kuten lainmukaisuutta, käyttötarkoitussidonnaisuutta ja tietojen minimointia. Rekisterinpitäjän vastuulla on pystyä osoittamaan, että periaatteita on noudatettu, mikä korostaa huolellisen suunnittelun ja dokumentoinnin tärkeyttä. Tekoälypohjaisten ennustemallien yhteydessä henkilötietojen käsittely edellyttää aina lainmukaista perustetta, ja usein soveltuvin vaihtoehto on rekisteröidyn suostumus. Suostumuksen on oltava annettu etukäteen, sen on oltava tietoinen, vapaaehtoinen, yksilöity ja yksiselitteinen, eli henkilön on ymmärrettävä, mihin ja miten hänen tietojensa käytetään, ja suostumus on ilmaistava selvästi esimerkiksi sanallisesti tai valitsemalla

aktiivisesti suostumuksen antava vaihtoehto. Jos käsitellään arkaluonteisia tietoja, kuten terveystietoja, suostumuksen on lisäksi oltava nimenomainen. Suostumus on voitava peruuttaa milloin tahansa, ja sen peruminen ei saa aiheuttaa henkilölle haittaa. Tärkeää on myös muistaa, että vaikka suostumus olisi saatu asianmukaisesti, se ei vapauta rekisterinpitäjää noudattamasta muita tietosuojaperiaatteita, kuten tietojen minimointia, käyttötarkoituksen rajaamista ja säilytysaikojen rajoittamista. Anonyymit tiedot jäävät GDPR:n ulkopuolelle, koska niistä on poistettu kaikki tunnistetiedot siten, ettei henkilöä voi enää tunnistaa. (Koivisto ym., 2023, ss. 71–72, 75–77; Regulation | 2016/679 | Gdpr | EUR-Lex, 2016; Sitra, 2022, ss. 36–37)

Tietosuojasetuksen rinnalle tullut EU:n tekoälyasetus (EU Artificial Intelligence Act 2024/1689) luokittelee AI-järjestelmät neljään eri riskitasoon, jotka ovat kielletty, korkean riskin, rajoitetun riskin ja vähäisen riskin järjestelmät. Tekoälyjärjestelmien valvonta ja virheiden tunnistaminen on varmistettava viranhaltijan osaamisen avulla. EU:n tekoälyasetuksen mukaan korkean riskin tekoälyjärjestelmät on suunniteltava ja kehitettävä siten, että ihmisten on mahdollista valvoa niiden toimintaa käytön aikana. Korkean riskin järjestelmät ovat kaikkein säännellyimpiä. Ne voivat liittyä esimerkiksi turvallisuuden jo valmiiksi säännellyissä tuotteissa tai olla itsenäisiä tekoälyjärjestelmiä, joiden toimintahäiriö tai väärinkäyttö voi uhata ihmisten terveyttä, turvallisuutta, perusoikeuksia tai ympäristöä. Rajoitetun riskin järjestelmät, kuten chatbotit, voivat sisältää manipulaation tai harhaanjohtamisen riskejä. Niiden on oltava läpinäkyviä, käyttäjän on tiedettävä olevansa vuorovaikutuksessa tekoälyn kanssa, ellei se ole muuten ilmeistä. Lisäksi esimerkiksi tekoälyn luomat sisällöt on merkittävä asianmukaisesti. Pienimmän eli vähäisen riskin järjestelmiä, kuten esimerkiksi roskapostisuodattimia, ei rajoiteta tai säännellä erityisesti. Niiden käyttöön suositellaan kuitenkin yleisiä periaatteita, kuten ihmisen valvontaa, syrjimättömyyttä ja oikeudenmukaisuutta. (European Commission, 2025; Regulation | 2024/1689 | AI Act | EUR-Lex, ei pvm.; Trail, 2024)

Tekoälyasetus tähtää yhteisten EU-tason tekoälyn sääntelyhiekkalaatikoiden (engl. regulatory sandboxes) perustamiseen tekoälyinnovaatioiden edistämiseksi. Sääntelyhiekkalaatikko on työkalu, joka mahdollistaa yrityksille uusien ja innovatiivisten tuotteiden, palvelujen tai liiketoimintamallien kokeilemisen valvotussa ja suljetussa ympäristössä viranomaisen ohjauksessa. Sääntelyhiekkalaatikon avulla voidaan testata sellaisiakin tuotteita, jotka eivät ole nykysäännösten mukaisia, tällöin organisaatio saa vapautuksen tietyistä säädöksistä voidakseen testata innovaatiota todellisessa

ympäristössä valvotusti. Testauksesta kertyvä tekninen data auttaa sääntelyviranomaista ymmärtämään paremmin innovaatioiden riskejä ja mahdollisuuksia ja arvioimaan, onko nykyinen sääntely tarkoituksenmukaista vai onko sitä muutettava. Tekoälyasetuksen luonnos pyrkii selkeyttämään uusien yleisten tekoälysäätöjen ja voimassa olevien tietosuojasäädösten yhteensovitusta. Jäsenvaltioiden tulee liittää kansalliset tietosuojaviranomaiset mukaan sandboxien toimintaan. Henkilötietojen jatkokäsittely olisi sallittua sääntelyhiekkalaatikossa, jos tekoälyratkaisua kehitetään yleistä etua varten, kuten esimerkiksi rikosten ennaltaehkäisyyn, kansanterveyteen tai ympäristönsuojeluun liittyen. EU:n asetus velvoittaa kaikkia jäsenmaita perustamaan vähintään yhden kansallisen sandboxin elokuuhun 2026 mennessä. (European Parliament, 2022, ss. 1–3)

Tekoälyasetuksen lisäksi säädetty datasäädös, eli Euroopan Unionin Data Act, vaikuttaa datan jakamiseen ja käyttöön toimijoiden välillä. Datasäädöksen tavoitteena on vahvistaa eri toimijoiden mahdollisuuksia jakaa ja hyödyntää dataa. Sen on tarkoitus parantaa tiedon saatavuutta ja edistää sen monipuolista käyttöä ja uudelleenkäyttöä. Samalla säädös korostaa yksityisyyden suojan ja tietoturvan merkitystä. Datasäädös laajentaa tietosuoja-asetusta datan jakamisen helpottamisen osalta ja voi poikkeustarpeissa mahdollistaa julkisen sektorin pääsyn yritysten hallussa olevaan dataan. (Kuntaliitto, 2025; Regulation | 2023/2854 | Data Act | EUR-Lex, 2023)

Suomessa julkishallinnon toimintaa ohjaa lainsäädäntö, joka määrittelee tiedon luovuttamisen ja käytön reunaehdot viranomaistoiminnassa. Lakisääteisiä tehtäviä hoitaessaan viranomainen voi käyttää vain sellaista tietoa, jonka keräämisen laki sallii kyseisen tehtävän suorittamiseksi asiakkaalta tai muista lähteistä. Käytännössä tietojen luovutus viranomaisille edellyttää tapauskohtaista tietolupaa. Lisäksi eri organisaatioilla on usein päällekkäistä tietoa, mutta tietojen yhteensovittaminen on haastavaa, koska tiedot on tallennettu erilaisin termein ja eri käyttötarkoituksen mukaisena. (”Suomalaisten data-avaruuksien tilannekuva”, 2024) Perustuslaissa on jo säädetty perusoikeuksista hyvään hallintoon ja oikeusturvaan. Hallintolaki täsmentää hyvän hallinnon kriteereitä ja siinä säädetään hyvän hallinnon perusteista, hallintoasian käsittelystä, selvittämisestä, kuulemisesta, sekä asian ratkaisemisesta ja tiedoksiannosta. Hallintolain oikeusperiaatteisiin kuuluu myös yhdenvertaisuus ja luottamuksensuoja, joiden on katsottu edellyttävän erityistä huomiota päätösaunomaatiota hyödynnettäessä. Automaattisesta päätöksenteosta ja siihen liittyvistä seikoista säädetään hallintolaissa ratkaisemisen

edellytysten sekä ilmoitusvelvollisuuden osalta. Päätöksestä on käytävä ilmi, jos se on ratkaistu automaattisesti. (*Hallintolaki | 434/2003 | Lainsäädäntö | Finlex*, ei pvm.)

Hallintolaissa säädetään erikseen perusteluvelvollisuudesta, joka tarkoittaa, että päätös tulee perustella ja perustelusta tulee käydä ilmi ratkaisuun vaikuttaneet seikat ja sovelletut säännökset. Perusteluista tulee pystyä ymmärtämään, miksi ratkaisuun on päädytty. Perusteluvelvollisuus on sitä laajempi, mitä enemmän ratkaisuun sisältyy harkinnanvaraa. Automaatiota hyödyntävien päätöksentekojärjestelmien käytössä tulisi myös kuvata miten ratkaisuun on päädytty. Tämän vuoksi perusteltavuuden ja läpinäkyvyyden kannalta, tulee käydä ilmi paitsi automatisaation käyttö, myös säännöt, algoritmit ja ylipäänsä tiedot ratkaisuun johtaneesta päättelyketjusta. Päätöksenteon valmistelun automatisoitujen taustaprosessienkin tulee olla hyvän hallinnon mukaisia. Viranhaltijan tulee ymmärtää tekoälyn tuottaman tiedon taustat ja lähteet sekä varmistettava, että lopputulos on oikeasuhtainen tavoiteltuihin tarkoituksiin nähden. Jos tekoäly vaikuttaa tosiasiallisesti yksilön oikeusasemaan, kyseessä voi olla julkinen hallintotehtävä. (Sitra, 2022, ss. 24–25)

Tekoälyn kouluttamisessa on tärkeää varmistaa, että järjestelmät tukevat yhdenvertaisuuden toteutumista eivätkä aiheuta syrjiviä vaikutuksia. Tämä edellyttää, että opetusaineisto on virheetöntä, kattavaa ja edustaa monipuolisesti eri ihmisryhmiä. Jos data on vinoutunutta tai puutteellista, voi tekoälyjärjestelmä vahvistaa ennakkoluuloja tai syrjiä esimerkiksi iän, sukupuolen, kielen, etnisen taustan tai asuinalueen perusteella. Syrjinnän riskit tulee tunnistaa jo algoritmien suunnitteluvaiheessa, ja järjestelmän toimintaa on valvottava säännöllisesti. Yhdenvertaisuuden näkökulmasta laadukas ja eettisesti rakennettu tekoäly vaatii myös vaikutusten ennakoarviointia ja jatkuvaa seurantaa käytännön toiminnassa. (Sitra, 2022, ss. 34–36)

Tietosuojavaikutusten arviointi eli DPIA on tärkeä vaihe silloin, kun henkilötietoja aiotaan käsitellä tavalla, joka voi todennäköisesti aiheuttaa merkittäviä riskejä ihmisten oikeuksille ja vapauksille. Arviointi on tehtävä ennen kuin käsittely aloitetaan, ja siinä täytyy ottaa huomioon muun muassa, kuinka laajaa käsittely on, mihin tarkoitukseen tietoja käytetään ja millaisessa kontekstissa niitä käsitellään. DPIA tulee erityisesti kyseeseen, jos henkilötietoja käytetään automaattiseen päätöksentekoon, joka voi vaikuttaa merkittävästi yksittäisen ihmisen elämään, tai jos käsitellään suuria määriä arkaluonteisia tietoja, kuten terveystietoja. Sosiaali- ja terveysalalla tällaisia tilanteita syntyy helposti esimerkiksi ennustemalleja kehittäessä ja käytettäessä, koska ne voivat vaikuttaa suoraan siihen, millaisia palveluja tai etuuksia ihminen saa. Vaikutustenarvioinnissa rekisterinpitäjän on

käytävä läpi, mitä henkilötietoja kerätään ja miksi, onko käsittely välttämätöntä ja kohtuullista tarkoitukseensa nähden, sekä millaisia riskejä yksittäiselle ihmiselle voi syntyä. Samalla täytyy kuvata ne keinot, joilla riskeihin varaudutaan ja varmistetaan, että tietosuojalainsäädäntöä noudatetaan. DPIA toimii myös keinona dokumentoida, että kaikki on tehty oikein, ja siksi sen laatimista suositellaan silloinkin, kun ei ole täysin varmaa, onko arviointi välttämätön. (Koivisto ym., 2023, ss. 80–81; Regulation | 2016/679 | Gdpr | EUR-Lex, 2016) Sosiaalialan ennustemallien kehittäminen ja hyödyntäminen edellyttää tyypillisesti asiakkaiden henkilötietojen käsittelyä. Ennustemallien käytön tulee olla tietosuojalainsäädännön mukaista, mikä edellyttää, että henkilötietojen käsittelylle olla hyväksyttävä peruste ja lainsäädännön vaatimukset täyttyvät. (Koivisto ym., 2023, ss. 71–72)

Suomessa tietosuojalaki täydentää yleistä tietosuoja-asetusta (GDPR). Se säätelee muun muassa erityisten henkilötietojen käsittelyä, tietosuojavaltuutetun tehtäviä ja poikkeuksia esimerkiksi tutkimuskäyttöön. Tietosuoja-asetus koskee kaikkia, jotka käsittelevät henkilötietoja, ja se jakaa vastuut kahdelle taholle, rekisterinpitäjille ja käsittelijöille. (Koivisto ym., 2023, s. 71; *Tietosuojalaki | 1050/2018 | Lainsäädäntö | Finlex*, 2018) Laki julkisen tiedon hallinnasta, eli ”Tiedonhallintalaki” täydentää muun muassa tietosuojalainsäädäntöä. Sen tavoitteena on parantaa viranomaisten välistä yhdenmukaista tiedonhallintaa, edistää tietojen yhteentoimivuutta ja varmistaa julkisen hallinnon tiedonhallinnan laatu, turvallisuus ja läpinäkyvyys. Laki koskee erityisesti viranomaisia ja muita julkista hallintotehtävää hoitavia tahoja. Tiedonhallintalaki velvoittaa viranomaisia dokumentoimaan tiedonhallintansa suunnitelmallisesti niin sanotun tiedonhallintamallin avulla. Siinä kuvataan muun muassa, miten tietoa käsitellään, säilytetään ja suojataan sekä millaisia tietovarantoja ja tietojärjestelmiä viranomaisella on käytössään. Laki korostaa myös velvollisuutta huolehtia tietoturvasta ja siitä, että henkilötietojen käsittely täyttää sekä tietosuojalainsäädännön että julkisuuslain vaatimukset. Tiedonhallintalaissa säädetään myös rekisteröidyn oikeuksista, kuten oikeudesta saada tietoa siitä, mihin hänen tietojaan käytetään ja kenelle niitä luovutetaan. Laki tukee myös tietojen toissijaista käyttöä, kuten tutkimus- ja kehittämistarkoituksia, mutta korostaa, että tällainen käyttö edellyttää aina selkeitä käsittelyperusteita ja asianmukaisia suojatoimia. (*Laki julkisen hallinnon tiedonhallinnasta | 906/2019 | Lainsäädäntö | Finlex*, ei pvm.)

Laki sosiaali- ja terveystietojen toissijaisesta käytöstä eli ”toisilaki” (L 552/2019) säätelee, miten sosiaali- ja terveystietoja voidaan käyttää muuhun kuin alkuperäiseen

tarkoitukseensa, kuten tutkimukseen, tietojohdamiseen tai viranomaisvalvontaan. Laki täydentää EU:n tietosuojasetusta ja tuli voimaan toukokuussa 2019. Keskeisiä käyttökohteita ovat erityisesti tieteellinen tutkimus ja tietojohdaminen. Tavoitteena on tehdä luvanhakuprosesseista sujuvampia ja nopeampia, sekä helpottaa tietojen yhdistämistä eri lähteistä. Findata, eli sosiaali- ja terveystietojen tietolupaviranomainen, vastaa tietojen kokoamisesta, käsittelystä ja luovutuksesta toissijaiseen käyttöön. Findata myös myöntää tietoluvat, ylläpitää turvallista käyttöympäristöä ja huolehtii siitä, että julkaistavat tulokset ovat anonyymeja. Lisäksi myös tietyt rekisterinpitäjät voivat myöntää lupia omiin aineistoihinsa. Ennustemallien kehittämistä voi tietosuojasetuksen mukaan joissakin tapauksissa perustella suostumuksella, mutta toisilaki voi monissa tilanteissa tarjota selkeämmän ja vakiintuneemman oikeudellisen perustan tällaiselle kehittämistyölle. (Koivisto ym., 2023, ss. 71–72; *Laki sosiaali- ja terveystietojen toissijaisesta käytöstä* | 552/2019 | *Suomen säädöskokoelma* | *Finlex*, 2019)

5 Opinnäytetyön toteuttaminen

Opinnäytetyö toteutettiin kartoittavana ja vertailevana kirjallisuuskatsauksena. Tässä työssä hyödynnetään integroivan kirjallisuuskatsauksen menetelmää, jonka avulla tarkastellaan ennustemallien käyttöä työllisyyspalveluissa aiemman tutkimustiedon, asiantuntija-artikkelien, blogikirjoitusten ja erilaisten raporttien pohjalta. Integroiva kirjallisuuskatsaus yhdistää monipuolista aineistoa, jotta aiheesta voidaan muodostaa kokonaisvaltainen kuva. Se mahdollistaa laaja-alaisen yleiskatsauksen tekemisen ilman järjestelmälliselle kirjallisuuskatsaukselle asetettuja tiukkoja metodisia vaatimuksia. Tällaisia vaatimuksia ovat esimerkiksi aineiston tarkka rajaaminen tai mukaan otettujen tutkimusten laadun arviointi. Menetelmän tavoitteena on jäsentää keskeisiä teemoja, tunnistaa aiemmissa tutkimuksissa esiin nousseita havaintoja sekä paikantaa aihealueita, jotka kaipaavat lisää tutkimusta. (Salminen, 2011, s. 8) Tarkoituksena on kerätä ja yhteensovittaa eri lähteistä mahdollisimman ajankohtaista tietoa, jotta saadaan kattava kuva teknologioiden käytöstä, tietoturvan toteutumisesta sekä eettisistä näkökulmista. Katsauksen kautta pyritään tunnistamaan myös niitä haasteita ja kehittämistarpeita, jotka liittyvät ennustemallien soveltamiseen palvelutarpeen arvioinnissa.

Tässä kirjallisuuskatsauksessa tarkastellaan, miten tekoälyyn pohjautuvia ennustemalleja voitaisi hyödyntää työllistymisen edistämässä ja monialaisen palvelutarpeen arvioinnissa,

erityisesti Suomessa käyttöön otetun TYM-yhteistoimintamallin näkökulmasta. Katsauksen tarkoituksena on arvioida, millaisia mahdollisuuksia ja rajoitteita ennustemallien käyttöön liittyy TYM-yhteistyömallissa, kun huomioon otetaan sekä toiminnalliset tavoitteet että lainsäädännölliset ja eettiset reunaehdot. Vertailun tueksi tarkastellaan valikoituja EU-maita, joissa on kehitetty vastaavanlaisia työllisyyden tuen malleja. Näiden maiden kautta pyritään ymmärtämään, miten tiedonkulku ja tiedonjako on järjestetty ennustemallien hyödyntämiseksi, esimerkiksi mitä tietoja voidaan jakaa eri toimijoiden välillä, millaisia teknisiä tai oikeudellisia ratkaisuja on käytössä, ja missä kulkee perusteltavuuden raja toisen viranomaisen jakaman tiedon käytössä. Vertailun avulla pyritään hahmottamaan, millä tavoin Suomen nykytila asemoituu suhteessa muihin maihin, eli missä ollaan nyt, mihin ollaan menossa, ja voidaanko muiden maiden kokemuksista oppia ennustemallien käyttöönoton hyödyntämiseen. Tarkasteltavat kysymykset ovat keskeisiä siksi, että ennustemallien toimivuus ja vaikuttavuus TYM-toiminnassa edellyttävät paitsi riittävää ja ajantasaista tietopohjaa myös selkeitä pelisääntöjä tiedon jakamisen, vastuukysymysten ja yhteisrekisterinpitäjyyden osalta. Monet kysymykset ovat ajankohtaisia mutta edelleen jossain määrin avoimia sekä Suomessa että muualla Euroopassa. Yleisesti ottaen kehittyneiden teknologioiden ja datan hyödyntäminen, mukaan lukien datan yhdistäminen on harvinaisempaa julkisella sektorilla kuin yksityisellä sektorilla, ja harvinaisempaa sosiaaliturvassa kuin esimerkiksi terveydenhuollossa. Ennustemallien käytöstä terveydenhuollossa löytyi runsaasti esimerkkejä, katsauksessa pyrittiin kuitenkin painottamaan vastaavaa ennustemallin käyttöä työllisyyden tukemisessa.

Tavoitteena on luoda yleiskuva siitä, millaisia ennustemalleja ja muita tunnistamisen työkaluja ulkomailla on kehitetty tai kehitteillä. Katsauksessa huomioidaan myös muiden maiden dataekosysteemit, tekniset toteutukset sekä oikeudelliset ja eettiset haasteet, joita on kohdattu. Tiedonhaku on toteutettu kirjallisuuskatsauksena. Katsaukseen on etsitty monipuolisia julkaisuja, kuten kartoituksia, tutkimuksia, raportteja ja katsauksia. Tätä katsausta varten haettiin artikkeleita ja kirjallisuutta kansainvälisistä tietokannoista sekä muita ajantasaisia ja soveltuvia lähteitä, kuten viranomaisten julkaisuja, politiikkaraportteja, asiantuntijalausuntoja ja media-aineistoja. Tiedonhakuun käytettiin myös hakukoneita, kuten Googlea, erityisesti silloin kun etsittiin ajankohtaisia tai vähemmän dokumentoituja käytännön esimerkkejä ja kokeiluja. Koska ennustemallien ja algoritmien käyttö julkisessa hallinnossa ei aina ole systemaattisesti raportoitu tieteellisessä kirjallisuudessa, oli tarpeen hyödyntää myös kansallisten virastojen tai kansainvälisten järjestöjen julkaisuja sekä mediassa esiin nousseita tapausesimerkkejä. Näin pyrittiin varmistamaan mahdollisimman

laaja ja ajantasainen kuva eri maiden ratkaisuksista ja kokemuksista. Lopullinen kirjallisuuskatsaus sisältää yli 50 läpikäytyä lähdettä. Lähdeaineistojen läpikäynnissä hyödynnettiin tekoälyä, jonka avulla tunnistettiin ja koottiin yhteen keskeiset sisällöt. Lähteiden läpikäymisessä käytettiin apuna myös Chrome selaimen Google Translatea kun lähteinä olivat muut kuin suomenkieliset sivustot. Englanninkielisessä tiivistelmässä hyödynnettiin kieliasun tarkastamisessa ja oikeinkirjoituksessa Grammarly kirjoittamisen avustajaa. Kaikki käännökset ja yhteenvedot tarkistettiin ja uudelleenkirjoitettiin manuaalisesti ennen niiden sisällyttämistä analyysiin.

6 Ennustemallien käyttötapaukset

Tässä luvussa tarkastellaan ennustemallien käyttötapauksia Suomessa ja valikoiduissa vertailumaissa. Käyttötapauksia etsittiin lähteistä nimenomaan työllistymisen ja työkyvyn ennakkointiin liittyvien ennustemallien käytössä. Tarkoituksena on tarkastella, miten ennustemalleja on kokeiltu ja käytetty muissa maissa ja luvun lopussa vertailla maita Suomen nykytilaan nähden. Suomen kannalta pääpaino on julkishallinnon esimerkeissä ja erityisesti TYM-yhteistoimintamallissa. Esimerkeissä on tarkoituksena keskittyä muiden maiden vastaavien toimijoiden ratkaisuihin datanjakamisen kulttuurin, sekä toiminnallisten ja oikeudellisten edellytysten näkökulmasta.

6.1 Käyttötapaukset Suomessa

Suomen digitalisaation nykytila on vahva ja kehittyy edelleen EU:n Digitaalinen vuosikymmen 2030-ohjelman tavoitteiden mukaisesti. Suomi erottuu edukseen erityisesti kansalaisten korkealla digiosaamisella ja sosiaalisella luottamuksella, ja sillä on EU:n parhaimpiin kuuluvat digitaaliset julkiset palvelut. Suomen digitaalista kehitystä ohjataan kansallisella digikompassilla, joka on laadittu ministerityöryhmän ohjauksessa ja digitoimiston valmistelussa. Kompassi perustuu EU:n Digitaalinen vuosikymmen 2030-ohjelmaan, ja se tarjoaa yhteisen suunnan digitalisaation ja datatalouden kehittämiseksi. Suomi edistää aktiivisesti myös eurooppalaisen digitaalisen identiteettilompakon kehittämistä sekä kansallisesti että EU-tason rajat ylittävissä kokeiluissa. Valtioneuvoston julkaiseman nelikenttäanalyysin mukaan Suomen vahvasta digitalisaatio-osaamisesta ja teknologiaympönteisyydestä huolimatta Suomi on jäämässä globaalista datatalouden kasvusta. Datatalouden, alustatalouden ja toimivien digitaalisten palveluiden kehitys

edellyttää sujuvaa datan jakamista eri toimijoiden kesken, mutta Suomessa tällaisia palveluja on syntynyt vain vähän. Tällä hetkellä datan jakaminen on vaikeaa, koska data jää yhden toimijan haltuun eikä kattavia ratkaisuja datan yhdistelyyn ole tarjolla. Hyötyjä saataisiin silloin, kun dataa olisi mahdollista yhdistää ja analysoida uusilla tavoilla esimerkiksi digitaalisten palvelujen, prosessien, tuotteiden tai liiketoimintamallien kehittämiseksi. Yritysten ketterä ja reilu pääsy dataan sekä rohkeus jakaa dataa edistäisi datatalouden kasvua. Suomen osalta julkishallinnon digitalisaatioon on keskitytty myös luvussa 4. (Valtioneuvosto, 2022, ss. 3–4, 13–14)

Suomessa on toteutettu useita ennustemallien kokeiluja sosiaali- ja terveysalalla, mutta ne ovat olleet pääosin yksittäisten toimijoiden omia hankkeita, joissa on hyödynnetty toimijakohtaista dataa. Lisäksi Suomessa toimitaan usein ns. silloissa, eli toimijat lähtevät liikkeelle organisatorisesta tarkastelukulmasta. Tässä kartoituksessa keskitytään uusimpiin ennustemallikokeiluihin ja käyttötapauksiin, jotka liittyvät työllistymisen ja työkyvyn tukemiseen. Työkyvyttömyyden ennustemallia on kokeiltu aiemmin Eläketurvakeskuksessa 2018 sekä Kelassa vuosina 2021 ja 2023. Eläketurvakeskuksen ennustemallikokeilussa hyödynnettiin ennustemallia työkyvyttömyyseläkkeelle joutumisen ennustamiseen ja Kelassa kehitetyn ennustemallin avulla voitaisi tulorekisteriä apuna käyttäen havaita työkyvyttömyysetuuden varaan jääminen. (Kelakanava, 2024)

Suomessa Eläketurvakeskuksen ennustemallikokeilu käytti jo vuonna 2018 koneoppimistekniikoita työkyvyttömyyseläkkeelle joutumisen ennustamiseen. Eläketurvakeskuksella on pääsy laajoihin rekisteriaineistoihin, kuten ansainta- ja eläkepäätosrekisteri, jotka kattavat lähestulkoon kaikki suomalaiset työntekijät ja eläkeläiset. Projektissa koneoppimismallia opetettiin ennustamaan työkyvyttömyyseläkkeen alkamista kahden vuoden päähän. Aineistona käytettiin 240 000 työkyvyttömyyseläkkeelle joutuneen henkilön anonymisoituja sosioekonomisia-, ansio- ja etuustietoja 2–10 vuotta ennen eläkkeen alkua. Verrokkeina oli 240 000 henkilöä, jotka eivät siirtyneet työkyvyttömyyseläkkeelle. Näin algoritmi pystyi oppimaan työkyvyttömyyttä ennakoivia tekijöitä. Henkilöitä yksilöivät tiedot poistettiin aineistosta ja aineiston käsittely tehtiin Eläketurvakeskuksen omassa käyttöympäristössä ilman pilvipalveluja. Kokeilu osoitti, että tietojen perusteella voidaan ennustaa työkyvyttömyyseläkkeen alkaminen melko tarkasti, sillä noin 78 % kahden vuoden sisällä alkavista eläkkeistä tunnistettiin oikein. Koneoppimistekniikat vaikuttavat potentiaaliselta apuvälineeltä asiantuntijatyön tueksi ja voivat tukea asiantuntijoita varhaisen tuen tarjoamisessa. (Eläketurvakeskus, 2018)

Kela ja Helsingin kaupunki toteutti syksyllä 2024 yhteisen työdataekosysteemikokeilun. Kokeilussa kehitettiin ratkaisuja osatyökykyisten ja työttömien palvelujen parantamiseksi. Kokeilun tarkoituksena oli lisätä osaamista dataekosysteemien hyödyntämisestä ja niissä toimimisessa, tuottaa yleisiä suosituksia ja konkreettisia toimenpiteitä sekä auttaa työllisyysenhoidon asiantuntijoita tunnistamaan asiakkaiden tarpeita yhdistämällä dataa eri viranomaistahoilta. Kokeilu koostui kahdesta toisiaan täydentävästä osasta, työttömän asiakkaan tilannekuva -prototyypistä, jonka avulla asiantuntijat pystyivät tarkastelemaan yhteisen datan hyödyntämismahdollisuuksia, sekä työkyvyn ennustemalli, jonka tavoitteena oli tuottaa ennakoivaa tietoa asiakkaan työkyvyn riskeistä. Kokeilun tuloksena ennustemallin avulla voidaan ”melko varmasti” havaita tukien varaan jääminen. Ennustemallia voidaan hyödyntää Kelan asiantuntijoille siten, että listataan suurimmassa riskissä olevat asiakkaat ja pyritään auttamaan heitä ohjaamalla oikeiden palveluiden piiriin. Mallin haasteena on, että monen asiakkaan näkökulmasta ennuste tulee liian myöhäisessä vaiheessa, kun tilanne on jo liian paha. (Mattila, 2024)

2024 tehdyssä ennustemallikokeilussa päivitettiin Kelassa jo aiemmin kehitettyä ennustemallia. Tietosuojavaikutusten arviointi tehtiin ennen ennustemallin tuottamista. Kokeilussa käytettiin pelkästään pseudonymisoitua aineistoa Kelan tietovarastosta. Kelalla on jo käytettävissä paljon asiakkaan tietoja, kuten sairauspäiväraha-, kuntoutus-, reseptilääkeosto-, tulorekisteri-, väestötietojärjestelmän ja ansiopäivärahan sekä työttömyysturvan maksutiedot. Dataan lisättiin uusia muuttujia ennustemallin taustalle, tämä paransi ennustemallin tuloksia. Kokeilussa testattiin kahta vastemuuttujaa, työkyvyttömyyseläkkeelle tai kuntoutustuelle päätyminen Eläketurvakeskuksen datan perusteella ja työkyvyttömyyseläkkeelle tai vastaavalle työkyvyttömyysetuudelle päätyminen tulorekisterin datan perusteella. Jos Kelan ja Helsingin kaupungin dataa olisi mahdollista yhdistää, voisi todennäköisesti ennustemallia kehittää vielä tarkemmaksi. (Kelakanava, 2024)

Suomessa on laajasti tahtotilaa edistää julkishallinnon digitaalisten palveluiden yhteentoimivuutta. Tiedon sujuva liikkuminen eri toimijoiden välillä on kuitenkin yhä haaste, ja palveluiden kehittämiseksi tarvitaan selkeämpi kuva siitä, mitä tietoa on saatavilla ja missä. Yhteistä koordinoivaa foorumia, joka ohjaisi suuntaa, teknisiä ratkaisuja ja yhteentoimivuutta tukevia kokeiluja, ei ole toistaiseksi ollut. Kelan vetämä epämuodollinen työryhmä on kehittänyt julkishallinnon data-avaruuksia, ja Virtual Finland -testialusta tarjoaa mahdollisuuden kokeilla data-avaruuksien ratkaisuja ja datan tuotteistamista. EU:ssa

julkishallinnon data-avaruus tunnetaan nimellä Data Space for Public Administration, mutta se ei ole vielä konkretisoitunut. (Poikola ym., 2024, Osa 4)

6.2 Käyttötapaukset vertailumaissa

Tässä osiossa esitellään kolme eri EU-maata, joissa on kokeiltu tai on edelleen käytössä työllistymiseen ja työkyvyn tukemiseen liittyviä ennustemalleja. Maiksi valikoitui Alankomaat, Tanska ja Viro, koska ne edustavat kolmea erilaista mutta kiinnostavaa lähestymistapaa työllistymisen ja työkyvyn ennakointiin liittyvien ennustemallien käytössä. Viro, Tanska ja Alankomaat jakavat Suomen kanssa samankaltaisen hyvinvointivaltion rakenteen sekä demografisia ja sosiaalipoliittisia piirteitä. Vertailumaissa on Suomen tavoin panostettu laaja-alaisesti työllisyyspalveluihin sekä varhaisen puuttumisen malleihin. Maat ovat Suomen näkökulmasta kiinnostavia, koska digitalisaatio ja tekoäly nähdään keskeisinä välineinä ennakoivan soten ja palveluohjauksen toteuttamisessa. Valikoidut maat tarjoavat erilaisia vertailukohtia ja näkökulmia sekä arvokkaita oppeja tilanteista, joita tahdotaan välttää ennustemallien suunnittelussa. Alankomaat on tunnettu kehittyneestä algoritmien hallintarakenteestaan ja pitkälle edenneestä datavetoisesta julkishallinnosta, mutta siellä on myös koettu merkittäviä epäonnistumisia, jotka tarjoavat arvokkaita oppimiskokemuksia vastuullisesta mallien käytöstä. Tanska puolestaan on kokeillut tekoälypohjaisia järjestelmiä, kuten ASTA-mallia, työllistymisriskiä arvioidessaan, ja tuo esiin haasteita mallien hyväksyttävyydessä ja läpinäkyvyydessä ammattilaisten näkökulmasta. Viro toimii esimerkkinä digitaalisen hallinnon edelläkävijästä, jossa tekoäly ja rekisteritieto on integroitu osaksi julkisia palveluja tavoilla, jotka mahdollistavat ennakoivan tuen tarjoamisen jo varhaisessa vaiheessa. Nämä kolme maata tarjoavat näin sekä teknologisesti että hallinnollisesti erilaisia konteksteja, joiden vertailu auttaa ymmärtämään, millaisia mahdollisuuksia ja reunaehtoja ennustemallien käyttöön liittyy eurooppalaisessa työllisyyspolitiikassa.

6.2.1 Alankomaat

Alankomaiden hallituksen tavoitteena on, että Alankomaat toimii digitaalisena edelläkävijänä. ”Nederland Digitaal” -strategia pyrkii vastaamaan digitalisaation nopeuteen ja digitalisaation yhteiskunnallisiin vaikutuksiin. Lähtökohtana on, että kaikki voivat osallistua digitaaliseen yhteiskuntaan turvallisesti, yksityisyyden suoja,

itseäänmääräämisoikeus ja digitaaliset taidot huomio huomioiden. (Zaken, 2018) Alankomaat johtaa Euroopan unionin tasolla vastuullisen tekoälyn käytössä julkisella sektorilla, ja se on vuonna 2024 sijoittunut ensimmäiseksi Global Index on Responsible AI -vertailussa. Maan kokemus on tärkeä esimerkki hyvästä hallintamallista, erityisesti siitä näkökulmasta kuinka kansalaiset hyväksyvät tekoälyjärjestelmien käyttöä päätöksenteossa. Alankomaiden keskeisimmät välineet tekoälyn käytön hallintaan ovat julkinen kansallinen rekisteri, jossa viranomaiset julkaisevat käyttämänsä algoritmit ja niin sanottu algoritmikehys, joka on ohjeistus ja työkalu julkishallinnolle vastuulliseen sekä eettiseen tekoälyn käyttöön. (Popa, 2025, Osa 1)

Alankomaissa palkansaajien työkyvyttömyyseläkejärjestelmästä vastaava julkinen työvoima- ja sosiaalivakuutuslaitos UWV (Uitvoeringsinstituut Werknemersverzekeringen) on käyttänyt koneoppimismalleja muun muassa työttömyystietojen analysoinnissa ja lyhyen aikavälin työmarkkinaennusteiden laatimisessa valtakunnallisesti. Ensimmäinen versio kehitettiin vuosina 2007–2010, otettiin alueellisesti käyttöön vuodesta 2011 ja koko maassa vuodesta 2015. Uusi versio otettiin käyttöön 2018. Heidän järjestelmänsä voivat tunnistaa työnhakijoiden riskiryhmät ja ennustaa työllistymismahdollisuuksia ja antaa suosituksia työllistymisen tueksi. Näiden työkalujen etuna on, että ne mahdollistavat resurssien tehokkaan ja oikeudenmukaisen kohdentamisen niille työnhakijoille, jotka tarvitsevat eniten tukea. Samalla palvelut voidaan kohdistaa tarkemmin heidän yksilöllisiin tarpeisiinsa. Uudistetun mallin myötä tiedonkeruu laajennettiin valtakunnalliseksi, malli kalibroitiin vastaamaan palvelurakenteiden muutoksia, kuten palvelujen digitalisoitumista ja tilastolliset menetelmät korvattiin kehittyneemmällä. Työkalun kehittämisessä ja toiminnassa hyödynnetään rekisteripohjaista dataa, jota kerää ja hallinnoi erityisesti UWV. UWV toimii sosiaaliasiain- ja työllisyysministeriön valvonnan alla ja ylläpitää sekä tietovarastoa että niihin liittyviä prosesseja itse, keskittyen lisäksi tietosuojaan ja julkisen sektorin vastuullisuuteen. (Havinga ym., 2023, ss. 110–111; Wijnhoven ym., 2023, ss. 110–112)

UWV:n Work Profiler ennustaa, kuinka todennäköisesti äskettäin työttömäksi jäänyt henkilö työllistyy vuoden sisällä. Ennuste auttaa kohdentamaan palveluja niille, joilla on työllistymisen haasteita tai esteitä. Työkalu yhdistää rekisteritietoa kuten ikä, koulutus, työhistoria, työttömyysetuudet ja itse täytettyä kyselydataa kuten motivaatio, terveydentila, osaaminen sekä työnhakukäyttäytyminen. Malli tukee asiantuntijoita työssä, mutta lopullisen päätöksen ja ohjauksen tekee ihminen. Work Profiler on edelleen käytössä Alankomaissa ja käyttö tukee työvoimapalveluita. Työkalun teoreettinen perusta auttaa

asiantuntijoita ymmärtämään, miksi tietyt tekijät vaikuttavat työllistymiseen ja miten niihin voidaan vaikuttaa. Ei muokattavia tekijöitä ovat esimerkiksi sukupuoli, ikä ja nykyinen koulutus. Muokattavia tekijöitä taas ovat motivaatio, oma terveystilanne ja työnhakuaktiivisuus. Tämän ansiosta ammattilaiset voivat suunnitella yksilöllisesti kohdennettuja palveluita, ja työnhakijat voivat ymmärtää omaa tilannettaan paremmin. Tämä lisää myös työkalun hyväksyttävyyttä ja luottamusta. Ammattilaisten hyväksyntä on keskeistä, jos malli ei vaikuta luotettavalta tai ei tuo lisäarvoa, sen tuottamaan tietoon ei luoteta tai sitä ei tahdota käyttää. Work Profiler yhdistää algoritmin objektiivisuuden, sekä asiantuntijan kokemuksen, jolloin virheen riski pienenee ja arvio on kokonaisuutena tarkempi kuin pelkkä algoritmi tai asiantuntijan työ yksinään. Työkalu on tarkoitettu ammattilaista tukevaksi, eikä se korvaa asiantuntijaa. Work Profilerin myötä huomattiin, että mallin päivittäminen on tärkeää, koska työmarkkinoiden tarpeet muuttuvat, yhä useammat palvelut ja palvelutilanteet digitalisoituvat ja taloussuhdanteet vaikuttavat siihen, miten monelle ja millaisia palveluja voidaan tarjota. (Wijnhoven ym., 2023; *Work Profiler (Werkverkenner)* | CEDEFOP, 2017, ss. 113–120)

Alankomaat lanseerasi joulukuussa 2022 nykyisen Algoritmirekisterin, jossa julkishallinto julkaisee tietoja käytössä olevista algoritmeista. Rekisterin tarkoitus on lisätä läpinäkyvyyttä ja luottamusta kansalaisten ja päättäjien keskuudessa. Kun valtio on avoin algoritmien käytöstä, kansalaiset, organisaatiot ja media voivat seurata ja valvoa, että lakeja ja sääntöjä noudatetaan. Alankomaiden algoritmirekisterissä on yli 500 algoritmia. Puolessatoista vuodessa rekisteriin on lisätty noin 400 viranomaisten käyttämää algoritmia, ja erityisesti kunnat ovat olleet edelläkävijöitä valtionhallintoon verrattuna. Tavoitteena on rekisteröidä kaikki vaikutukseltaan merkittävät, ns. ”korkean riskin” algoritmit vuoden 2025 loppuun mennessä. (Ibestuur, 2024) Jokaisesta algoritmista julkaistaan mm. käyttötarkoitus, lainsäädännöllinen perusta, käytetyt tietolähteet, riskit ja ehkäisytimet. Julkishallinnon algoritmien omistaja vastaa, että algoritmi on turvallinen, eettinen ja perustuu voimassa olevaan sääntelyyn. Autoriteit Persoonsgegevens (AP), eli Alankomaan tietosuojaviranomainen koordinoi valvontaa eli noudattavatko julkishallinnon algoritmit säädettyjä sääntöjä. GDPR on ollut keskeinen sääntelykehys myös ennustemallien käytössä työllisyyspalveluissa. Työllisyyspalveluissa algoritmeja on käytetty työnhakijoiden profilointiin. AP katsoi tällaisen seurannan GDPR-kyseenalaiseksi ja vaati korjaavia toimia. Alankomaat noudattaa tarkasti EU:n uuden AI-asetuksen mukaista riskiperustaista mallia, jossa työnhakuun, työntekijöihin ja palveluihin liittyvät koneoppimissovellukset ovat korkean riskin järjestelmiä. Alankomaissa kaikki automaattiseen päätöksentekoon tai profilointiin,

kuten UWV:n Work Profilerissa, perustuva datan käsittely luokitellaan tiukkaan GDPR-sääntelyyn. Tämän seurauksena DPIA (data protection impact assessment) on pakollinen kaikille korkean riskin järjestelmille, kuten koneoppimiseen työllisyyden ennustamiseksi. Mallinnuksen perusteet ja vaikutukset dokumentoitava tarkasti. (RvIG, 2025)

Alankomaissa algoritmien valvonta julkisella sektorilla on sekä viranomaisesti ohjattua että laajasti institutionalisoitua. Rekisteröintivelvoitteiden, eettisten kehysmallien ja riippumattomien auditointien avulla pyritään varmistamaan, että algoritmeja käytetään läpinäkyvästi, oikeudenmukaisesti ja kansalaisten luottamusta vaalien. Algoritmien valvonnan lisäksi viranomaiset vastaavat myös eri säädösten koordinoinnista. Alankomaat on nimennyt Autoriteit Consument & Markt (ACM) eli kuluttaja- ja markkinaviranomaisen kansalliseksi EU:n Data Actin datakoordinaattoriksi. ACM toimii päävastuun kantavana valvojana ja koordinoi toimimista muiden valvontaviranomaisten kuten tietosuojaviranomaisen kanssa. (Mous, 2024)

CBS (Centraal Bureau voor de Statistiek) eli Alankomaiden tilastokeskus voi lakisääteisesti yhdistää eri viranomaisten hallinnoimaa tietoa. Tilastokeskus (CBS) kerää hallinnollisia tietoja eri rekistereistä, jotka sisältävät tietoa henkilöistä, kotitalouksista, työpaikoista, yrityksistä, asunnoista ja muista vastaavista. Näissä aineistoissa kaikki henkilötunnisteet korvataan anonyymeillä linkitettävillä tunnuksilla, kuten anonyymillä henkilö-, kotitalous- tai osoitetunnuksella. CBS:n mikrodata tarjoaa tutkijoille ja organisaatioille mahdollisuuden tehdä tilastollista tutkimusta anonymisoidusta henkilötiedoista tiukkojen tietosuojavaatimusten alla. Pääsy aineistoon edellyttää organisaation valtuutusta ja projektihakemusta, joka arvioidaan GDPR:n ja toteuttamiskelpoisuuden näkökulmasta. CBS luo tutkimukselle suojatun työympäristön, jossa aineistoja voi analysoida. Tutkimustulokset tarkistetaan ennen julkaisua tietosuojariskien varalta, ja ne tulee julkaista julkisesti lähdeviittein. Esimerkiksi sosiaali- ja työministeriö (SZW), työllisyysviranomainen UWV sekä suunnittelutoimistot kuten Centraal Planbureau (CPB) ja Sociaal en Cultureel Planbureau (SCP) hyödyntävät mikrodataa kehittäessään pitkän aikavälin työvoimaennusteita ja arvioidessaan sosiaalipoliittisten toimenpiteiden vaikutuksia. CBS:n mikrodata on anonymisoitua ja tilastolliseen tutkimukseen tarkoitettua, eikä sitä siis käytetä yksittäisten kansalaisten asioiden käsittelyssä, kuten etuuspäätöksissä tai palveluohjauksessa. (Netherlands, 2025; Sivak & Stulp, 2024)

Alankomaissa on ollut epäonnistuneitakin ennustemallikokeiluja, jotka ovat herättäneet epäluottamusta ja aiheuttaneet kohuja. Alankomaiden SyRI-järjestelmä (System Risk Indication) otettiin käyttöön vuonna 2014 viranomaisten työkaluksi sosiaaliturva- ja hyvinvointijärjestelmään kohdistuvan petoksen torjuntaan. Se yhdisteli laajasti kansalaisten henkilötietoja useista rekistereistä, kuten asuminen, tulot, sakot ja sosiaalietuudet, tunnistaakseen poikkeavia kansalaisprofiileja, joita tulisi tutkia tarkemmin. Järjestelmä toimi salaisella algoritmilla, eikä kansalaisilla ollut pääsyä tietoihin siitä, miksi heidät mahdollisesti luokiteltiin riskitapaukseksi. Järjestelmää käytettiin pääosin matalatuloisilla alueilla, mikä johti syrjiviksi tulkittaviin käytäntöihin. Kansalaisjärjestöt ja tutkijat kritisoivat järjestelmää sen läpinäkymättömyydestä, mahdollisista vääristä positiivisista osumista ja heikosta oikeusturvasta. Haagin tuomioistuin kielsi SyRI:n käytön helmikuussa 2020 vedoten siihen, että järjestelmä rikkoi Euroopan ihmisoikeussopimuksen 8 artiklaa (yksityiselämän suoja). Tuomioistuimen mukaan SyRI ei täyttänyt vaatimusta oikeudenmukaisesta tasapainosta yksityisyyden suojan ja yhteiskunnallisten etujen välillä. Päätös pakotti julkisen hallinnon arvioimaan algoritmipohjaisten valvontatyökalujen käyttöä laajemmin. Kritiikin ytimessä oli ajatus siitä, että valtion tulisi rakentaa datankäyttönsä luottamuksen, läpinäkyvyyden ja oikeudenmukaisuuden periaatteille, ei massavalvonnan tai syrjivien algoritmien varaan. (AlgorithmWatch, 2020; Spisak, 2024, ss. 14–15)

6.2.2 Tanska

Vuonna 2021 Tanska sijoittui ensimmäiseksi Euroopan komission vuosittaisessa DESI-vertailussa (Digital Economy and Society Index), joka mittaa EU-maiden digitaalista toimintakykyä ja sen kehitystä. Tanska sai erityisen korkeat pisteet sähköisen hallinnon käytöstä ja avoimen datan hyödyntämisestä, joissa molemmissa se oli EU:n kärkimaa. (Koivisto ym., 2023, s. 71) Tanskan digitalisaatiostrategia vuosille 2022–2026 kokoaa yhteen toimenpiteitä, joilla pyritään parantamaan arkea digitaalisin keinoin. Tavoitteena on esimerkiksi parempi kyberturvallisuus, sujuvammat julkiset palvelut, digitaalinen terveydenhuolto ja vihreän siirtymän tukeminen teknologian avulla. Strategia korostaa myös vastuullisuutta, yksityisyyden suojaa ja eettistä digikehitystä. Keskiössä on ajatus, että digitaaliset ratkaisut palvelevat kansalaisia, yrityksiä ja yhteiskuntaa laajemmin. Julkisen ja yksityisen sektorin yhteistyöllä halutaan varmistaa, että kehitys etenee ihmisten tarpeet edellä. Tanskan hallitus korostaa, että julkiset IT-järjestelmät ovat kriittistä infrastruktuuria, aivan kuten tiet ja sillat, ja niitä on huollettava ja uudistettava säännöllisesti. Tavoitteena on rakentaa vahva digitaalinen perusta, jossa yhteiset ratkaisut,

kuten julkiset tietovarannot, ovat keskeisessä roolissa. Tavoitteena on, että vuoteen 2030 mennessä tanskalaiset ovat entistä valmiimpia tarttumaan digitalisaation tuomiin mahdollisuuksiin. (The Danish Government, 2022, ss. 62–70) Osana Tanskan digitalisaatiostrategiaa Tanskan digitaalinen virasto on yhteistyössä Tanskan paikallishallinnon ja Tanskan alueiden kanssa ottanut käyttöön digitaalisen suostumuksen. Viranomaiset voivat pyytää kansalaisten suostumusta turvallisesti, digitaalisesti ja yhdenmukaisesti. Kansalaislähtöisessä käyttöliittymässä kansalaiset voivat antaa suostumuksia, tarkastella tietojaan ja myös peruuttaa suostumuksen. Julkinen sektori pyytää yleensä monissa eri järjestelmissä lupaa käyttää ja jakaa henkilötietoja, jotka koskevat esimerkiksi terveyttä tai etuuksia. Tulevaisuudessa digitaalinen prosessi säästää arvokasta aikaa, jota voidaan käyttää asian käsittelyyn. (Digitaliserings-styrelsen, 2025)

Tanskan työmarkkinaviranomainen STAR (Styrelsen for Arbejdsmarked og Rekruttering) vastaa työllisyyspolitiikan toimeenpanosta ja kunnallisten työvoimapalveluiden (Jobcentereiden) toiminnan tukemisesta. Lisäksi STAR palvelee suuria työnantajia, toimii heidän yksittäisenä yhteyspisteenään sekä tarjoaa IT-infrastruktuurin muiden työvoimapalveluiden käyttöön. Yksi sen keskeisistä kehittämistoimista on ollut ennakoivien mallien ja algoritmien käyttö työvoimapalveluissa, erityisesti asiakkaiden ohjauksessa oikeisiin palveluihin. Malli yhdistää hallinnollista dataa sekä käyttäytymistietoja keräävää verkkokyselyä. Järjestelmä on vapaaehtoinen työnhakijoille, ja se tukee sosiaalityöntekijöiden päätöksentekoa, mutta ei automaattisesti ohjaa hakijoita aktiivimarkkinatoimiin. Nykyisistä käytössä olevista malleista ei löydy tarkempaa tietoa. (Bray & Gray, 2022, s. 26; OECD, 2025, s. 16)

Tanskan työmarkkinahallinnon yhteydessä kehitetty ASTA (Algoritmisk Sagsbehandlingsværktøj til Arbejdsmarkedet) oli algoritmiohjattu riskiprofilointityökalu, jonka tarkoituksena oli arvioida työttömien työnhakijoiden riskiä joutua pitkäaikaistyöttömiksi. Työkalu perustui koneoppimismalleihin, jotka käyttivät useita kymmeniä taustamuuttujia. Taustamuuttujina käytettiin henkilön ikää, koulutustaustaa, työhistoriaa ja maahanmuuttajataustaa, jolla laskettiin todennäköisyys, jolla työnhakija jää ilman työtä pitkällä aikavälillä. Mallin tuottama arvio luokitteli henkilön kolmeen riskiryhmään, vihreä matalan riskin, keltainen keskitason riskin tai punainen korkean riskin. (Flügge ym., 2022, s. 2)

ASTA oli suunniteltu päätöksenteon tueksi, ei automaattisen päätöksenteon välineeksi. Käytännössä sen oli tarkoitus auttaa kunnan työllisyysneuvonantajia suuntaamaan resursseja niille, joilla oli korkein riski jäädä ilman työtä. Työkalua pilotoitiin useissa kunnissa vuosina 2019–2022. Kokemukset vaihtelivat, osa kunnista koki algoritmin hyödylliseksi, mutta monissa paikoissa ilmeni vakavia epäluottamuksen ja läpinäkyvyyden ongelmia. Työntekijät kokivat, että ASTA:n tekemä luokittelu matalan, keskitason ja korkean riskin ryhmiin ei vastaa heidän omia kokemuksiaan siitä, kuinka moni asiakas työllistyy nopeasti. Useat sosiaalityöntekijät ilmoittivat, etteivät ymmärtäneet, miten algoritmin tuottamat riskiluokitukset muodostuivat. Tämä vaikeutti työkalun hyödyntämistä käytännössä. (Flügge ym., 2022, s. 3) Lisäksi ASTA herätti merkittäviä huolia tietosuojan ja syrjinnän näkökulmista. Tanskan tietosuojaviranomainen Datatilsynet arvioi, ettei työkalun käyttö täyttänyt GDPR:n edellyttämää tietojen minimoinnin periaatetta. Vuonna 2022 Tanskan työministeriö päätti lopettaa työkalun käytön vedoten teknisiin rajoitteisiin, oikeudellisiin epäselvyyksiin ja kunnallisten käyttäjien palautteeseen. Tanskan tietosuojaviranomainen katsoi, ettei kansalaisen suostumus voi toimia henkilötietojen käsittelyn oikeusperusteena yleisen tietosuojasetuksen (GDPR) mukaisesti, sillä tässä asiayhteydessä kansalaisen suostumusta ei voida pitää vapaaehtoisena. (Datatilsynet, 2022)

Tanskassa Tietosuojaviranomainen ja Kaupallinen valvontaviranomainen tulevat saamaan roolia Data Actin valvonnassa, mutta tarkat roolit ja työnjako ei ole vielä täysin selvä. Kansallinen lainsäädäntö tullaan päivittämään tukemaan EU:n vaatimuksia, mukaan lukien velvoitteet läpinäkyvyydestä ja vastuullisesta datan käytöstä. Tanskan tietosuojaviranomainen (Datatilsynet) valvoo GDPR:n noudattamista ja henkilötietojen käsittelyä ja osallistuu Data Actin soveltamisen valvontaan erityisesti yksityisyydensuojan näkökulmasta. Kilpailuvirasto (Konkurrence- og Forbrugerstyrelsen) on mukana valvomassa kilpailuun ja datan saatavuuteen liittyviä sääntöjä. ASTA:sta oppineena Tietosuojaviranomainen on korostanut, että henkilötietojen minimointi ja tietojen asianmukainen käsittely ovat välttämättömiä. Tämä on johtanut varovaisuuteen ja selkeisiin sääntöihin datan keruusta, analysoinnista ja ennustemallien käytöstä etenkin julkisissa palveluissa. Data Actin toimeenpanossa pyritään vahvistamaan selkeät vastuut ja avoimuus, jotta vastaavat ongelmat voidaan välttää. (Agency for Digital Government, 2025; Datatilsynet, 2025) Tanskassa Data Protection Act eli Databeskyttelsesloven on kansallinen laki, joka täydentää ja täsmentää EU:n yleistä tietosuojasetusta (GDPR).

Tanskassa on 2018 otettu käyttöön digitaalivalmista lainsäädäntöä koskevia periaatteita ja käytäntöjä. On olemassa useita esimerkkejä julkisista IT-hankkeista, jotka ovat merkittävästi kallistuneet ja viivästyneet, koska lainsäädäntö on muotoiltu ilman riittävää huomiointia myöhemmästä toimeenpanosta. Kansalaiset ja yritykset kohtaavat joillain aloilla monimutkaista sääntelyä, ja asioiden käsittely edellyttää usein ammatillista harkintaa. Tämä voi aiheuttaa epäselvyyttä siitä, mitä lakeja ja ohjeita tietyssä tilanteessa sovelletaan. Monimutkainen lainsäädäntö poikkeuksineen, epämääräiset termit ja lukuisat menettelyvaatimukset voivat vaikeuttaa tehokasta ja digitaalista hallintoa. Vuonna 2018 hallitus ja laaja poliittinen rintama solmivat sopimuksen, jonka tavoitteena oli tehdä lainsäädännöstä selkeämpää, yksinkertaisempaa ja digitaaliseen hallintointiin soveltuvaa. Tanskassa tämä tarkoittaa, että uutta lainsäädäntöä laadittaessa arvioidaan jo varhaisessa vaiheessa, miten se voidaan toteuttaa digitaalisesti, ja pyritään välttämään monimutkaisia poikkeuksia ja epäselviä käsitteitä. Myös vaikutukset julkisen hallinnon digitoteutukseen otetaan huomioon. Tanskassa digitaalivalmiuden edistäminen lainsäädännössä on osa laajempaa julkisen sektorin digitalisaatiostrategiaa, ja siihen liittyviä käytäntöjä on kehitetty ja juurrutettu vuosien aikana. Tanskassa on valmisteilla työllisyysuudistus, joka tuo kunnille oikeusperustan tekoälyn käytössä työllisyystoimissa. Uudistuksen ensimmäisten osien odotetaan tulevan voimaan helmikuussa 2026 ja viimeisten osien tammikuussa 2027. Tämä voi esimerkiksi koskea päätöksentekotukea, jotta tekoälyä voidaan käyttää resurssien vapauttamiseen asiankäsittelyssä. Työkaluna tekoälyllä on potentiaalia tehostaa työntekijöiden, mikä vapauttaa kansalaisten aikaa ja resursseja. (Agency for Digital Government, 2018, ss. 1–2; Styrelsen for Arbejdsmarked og Rekruttering, 2025)

6.2.3 Viro

Viron nopea teknologinen kehitys alkoi pian Neuvostoliitosta irtautumisen jälkeen. Vuonna 2001 käynnistetty e-Estonia-hanke loi digitaalisen ekosysteemin, jossa 99 % valtion palveluista on saatavilla verkossa. Verotuksesta 95 % ja pankkisiirroista 99 % hoidetaan sähköisesti. Tämä alusta mahdollistaa Viron hallitukselle digitaalisten innovaatioiden kokeilun muita maita nopeammin. Tällä hetkellä kehitteillä on tekoälysovelluksia julkisiin palveluihin, kuten työllisyyspalveluihin. Yksi ratkaisusta on koneoppiva päätöksentekotyökalu, joka tukee työttömyystoimistojen virkailijoita. Se analysoi yli 100 000 asiakastietoa ja ehdottaa parhaat keinot tukea työnhakijoita, kuten koulutus tai kielikurssi. Työkalu arvioi myös todennäköisyyttä työllistyä ja auttaa jakamaan työkuormaa. (Lago, 2021)

Viron työvoimapalvelu ja työttömyysvakuutusrahasto (EUIF) yhteistyössä IT-vaikutusten tutkimuskeskuksen (CITIS), Restan ja Nortalin kanssa on kehittänyt tekoälypohjaisen päätöksentekotyökalun nimeltä OTT, joka ennustaa työttömien henkilöiden työllistymismahdollisuuksia ja työttömyyden uusiutumisen riskiä seuraavien 180 päivän aikana. Tämä työkalu on suunniteltu tukemaan EUIF:n asiantuntijoita tarjoamalla heille asiakkaista tietoa mihin heidän kannattaa keskittyä ja mitkä asiat ovat prioriteetteja, auttaen heitä tekemään parempia päätöksiä resurssien kohdentamisessa niille, jotka niitä eniten tarvitsevat. OTT auttaa asiantuntijoita kohdentamaan avun sinne, missä sitä eniten tarvitaan. Henkilö, jolla on matala riski, löytää työtä todennäköisesti itsenäisesti, kun taas korkean riskin asiakkaat saavat enemmän tukea ja yksilöllisen suunnitelman. Alun perin työkalun oli tarkoitus ennustaa pitkäaikaistyöttömyyden riskiä, mutta käytännön kokemusten perusteella sen ennustettavuus parani, kun se keskittyi työllistymismahdollisuuksiin. OTT-työkalu auttaa EUIF:n neuvonantajia ymmärtämään paremmin asiakkaidensa tilanteita laskemalla todennäköisyyden uuden työn löytymiselle sekä riskin työttömäksi joutumiselle uudelleen. Tarkemmin, automatisoitu päätöksenteon tukityökalu antaa EUIF:n neuvonantajille taustatietoa arvioimalla todennäköistä aikaa, jolloin asiakas työllistyy. Tämä arvio perustuu työttömiin liittyviin tietoihin kyseisen työllisyysalueen nykyisestä työmarkkinatilanteesta ottaen huomioon esimerkiksi koulutuksen, asuinpaikan ja koulutustaustan. Lisäksi OTT tarjoaa tietoa siitä, mitkä tekijät vaikuttavat eniten riskipisteisiin ja paljastaa mahdollisia työttömyyden syitä. (Nortal, 2022b; Vihalemm ym., 2025, s. 1) OTT käyttää tekoälyssä random forest-mallia, joka on opetettu ja testattu viimeisen viiden vuoden työttömyystiedoilla. Malli arvioi kunkin työttömän henkilön työllistymismahdollisuudet käyttäen noin 60 eri attribuuttia, jotka liittyvät henkilöön, esim. koulutus, työkokemus, oikeus etuuksiin, terveydentila, sekä työmarkkinoihin, esim. vapaat työpaikat eri alueilla, uusien työttömien määrä. OTT-projekti on osoittanut, että modernit koneoppimismallit voivat parantaa julkisen sektorin prosesseja ja tehdä työstä tehokkaampaa ja miellyttävämpää, samalla kun kustannuksia säästyy. (Nortal, 2021)

Virossa lähes kaikki kansalaisten ja valtion välinen asiointi on digitaalista, äänestämisestä veroilmoituksiin ja pysäköintimaksuihin. Näiden palveluiden takana on X-Road, joka on infrastruktuuri digitaaliseen viestintään viranomaisten, kansalaisten ja yritysten välillä. Ohjelmisto yhdessä standardien ja lakien kanssa muodostaa Viron digitaalisen selkärangan. X-Roadin ydin on yksilölliset avaimet, jokaisella viranomaisella, kansalaisella ja yrityksellä on digitaalinen identiteetti. Tietoja vaihdetaan kahden osapuolen välillä suoraan, ilman kolmansien osapuolien palvelimia. Lakipykälät kieltävät tietojen turhan

kopioinnin. Jokainen viranomainen näkee ja hallinnoi vain omaa tietokantaansa, eikä tietoja kopioida turhaan tai säilytetä useissa eri paikoissa. (Nortal, 2022a)

Virossa suuri osa julkisista palveluista toimii verkossa, ja digitaaliset työkalut ovat keskeinen osa valtion arkea, mukana ovat muun muassa Viron työvoimapalvelu ja työttömyysvakuutusrahasto (EUIF). Vaikka EUIF käyttää työllistymispalveluissa myös ulkopuolisia palveluntarjoajia, tässä Viron esimerkissä keskitytään erityisesti siihen, miten eri valtion rekisterit vaihtavat tietoa automaattisesti keskenään. Viron digitaalinen infrastruktuuri eroaa paljon monien muiden maiden vastaavista, joten tämän mallin suora kopioiminen ei välttämättä toimi kaikkialla. Silti se antaa hyvän kuvan siitä, miten tiivis järjestelmien integrointi voi parantaa monia työllistymispalvelujen osa-alueita. Viron hallituksen digitalisaatiotavoitteet perustuvat useisiin periaatteisiin. Keskitettyjä tietokantoja ei ole, vaan julkiset organisaatiot sekä yritykset voivat kehittää digitaalisia järjestelmiään itsenäisesti. Turvallinen tiedonvaihto on varmistettu järjestelmäelementtien välillä. Tietojen eheys taataan käyttämällä KSI-lohkoketjuteknologiaa (Viron suunnittelema lohkoketjuteknologia) tiedonvaihdossa. KSI on avoin alusta, joten mikä tahansa organisaatio voi käyttää infrastruktuuria avoimen lähdekoodin ratkaisuna. Jatkuvat teknologiainvestoinnit ja säädösmuutokset päivittävät lainsäädäntöä, jotta datan käyttö sujuu yksityisyyttä suojaan. Oleellinen osa periaatteita käyttäjän näkökulmasta on, että yksilön tarvitsee toimittaa viranomaisille tarpeelliset tiedot vain kerran. Tiedot jaetaan sitten julkisten organisaatioiden kesken ja vaihdetaan turvallisesti niiden kesken, jotka tarvitsevat niitä palvelujen tuottamiseen. Kaikilla kansalaisilla on myös oikeus nähdä itseään koskevat tiedot, joita hallinnollisissa rekistereissä on kerätty, sekä miten julkiset organisaatiot käyttävät näitä tietoja. (Vodopivec, 2023, ss. 41–42)

X-road:n avulla EUIF pystyy keräämään suuren osan työnhakijoiden tiedoista eri rekistereistä sen sijaan, että tiedot kerättäisiin suoraan työnhakijoilta. Jos tiedot ovat jo olemassa jossain muussa rekisterissä, ne jaetaan EUIF:lle tarpeen mukaan palveluiden tarjoamista varten, eli niiden EUIF:n tehtävien ja tietotarpeiden mukaisesti, jotka on määritelty lainsäädännössä. Tähän asti Viron työttömyysvakuutusrahaston (EUIF) ja yksityisten palveluntarjoajien välinen yhteistyö on ollut vähäistä, koska tiedon suojaaminen ja tietoturva vaatimukset tekevät järjestelmään pääsyn monimutkaiseksi ja kalliiksi. (Vodopivec, 2023, ss. 43–44) Jokaisen organisaation, oli se sitten julkinen virasto tai yksityinen yritys, joka haluaa vaihtaa tietoja X-Roadin kautta, täytyy ensin liittyä X-Roadin jäseneksi hakemalla ja rekisteröitymällä Viron IT-viranomaiselle RIA:lle (Riigi Infosüsteemi

Amet). Tämän jälkeen jäsenen on asennettava turvapalvelin. Turvapalvelimen tehtävänä on salata ja purkaa X-Road-jäsenten välillä kulkevia tiedonvaihtoviestejä. Turvapalvelin hallinnoi käyttöoikeuksia, eli kukin jäsen voi itse määrittää, mitkä muut jäsenet pääsevät käyttämään sen palveluita. Säädösten mukaan jokainen jäsen ilmoittaa järjestelmänsä ja palvelunsa RIA:lle ennen käyttöönottoa, mikä mahdollistaa palveluiden yhdenmukaistamisen. Kaikki kansalaiset voivat nähdä heistä käydyt viestit, mikä lisää läpinäkyvyyttä ja auttaa havaitsemaan virheet tai väärinkäytökset heti. (Bharosa ym., 2020, Osa 5)

6.3 Käyttötapausten vertailu

Tässä vertailussa tarkastellaan Suomen lisäksi valikoituneita maita, Tanskaa, Alankomaita ja Viroa, jotka ovat kehittyneet eri tavoin julkisen hallinnon digitalisaatiossa. Vertailu on tehty tässä luvussa esitettyyn kirjallisuuskatsaukseen pohjautuen ja vertailussa peilataan myös aiemmissa teorialuvuissa esiintuotuja seikkoja. Tässä tarkastelussa kiinnitetään huomiota erityisesti siihen, millaisilla teknologisilla, lainsäädännöllisillä ja organisatorisilla ratkaisuilla nämä maat ovat onnistuneet yhdistämään tiedon hyödyntämisen, ennustemallit ja asiakaspalvelun tukemisen. Suomen näkökulmasta vertaillaan, miten nämä käytännöt eroavat suomalaisesta järjestelmästä ja mitä opittavaa niistä voisi löytyä. Vertailtavat maat jakavat yhteisen tavoitteen olla edelläkävijöitä digitalisaation ja tekoälyn hyödyntämisessä julkishallinnossa. Kehitystä ovat tukeneet erityisesti johdonmukaiset kansalliset strategiat, digitalisaatiota mahdollistavat lainsäädäntömuutokset sekä tekoälyn käyttöä ohjeistavien toimielinten perustaminen. Tässä suhteessa on tunnistettavissa useita yhtymäkohtia myös Suomen lähestymistapaan, erityisesti strategisella tasolla. Suomi oli yksi ensimmäisistä EU-maista, joka julkaisi oman kansallisen tekoälystrategiansa.

Taulukkoon (

Taulukko 1) on kerätty tiivistetysti pääkohdat tässä luvussa käsiteltävistä eroista muiden maiden ja Suomen tämänhetkisessä tilanteessa. Taulukossa on huomioitu mallien käyttö palveluohjauksessa, tiedonvaihto ja datanjakamisen kulttuuri, lainsäädännön tulkinta, organisointi ja mahdolliset eettiset haasteet. Taulukossa on tiivistetty muiden maiden esiin nousseet yhteiset piirteet samaan sarakkeeseen ja niitä verrataan Suomen tilanteeseen. Taulukon tarkoitus on antaa tiivis kokonaiskuva, alaluvuissa on eroteltu tarkemmalla tasolla

muiden maiden eroja ja yhteneväisyyksiä. Taulukon yhteenvedossa näkyvät pääkohdat ja maiden väliset erot on selitetty tarkemmin auki tulevissa alaluvuissa.

Taulukko 1. Käyttötapausten vertailu

	Muut maat	Suomi/TYM-malli	Huomiot ja opit
Ennustemallit palveluohjauksessa	Käytetään palveluohjauksen tueksi työllisyyspalveluissa	Ei tällä hetkellä käytössä palveluohjaukseen, kokeiluista lupaavia tuloksia	Mallien tuki voisi tehostaa työtä, mutta ei korvata täysin asiantuntijan harkintaa
Tiedonvaihto ja datanjakamisen kulttuuri	Keskitetyt alustat (X-Road, algoritmirekisterit)	Hajautunut, kullakin organisaatiolla oma alusta ja rekisteri	Tiedonvaihtorakenteet ovat este ennustemallien hyödyntämiselle. Yhteinen digitaalinen alusta voisi tukea TYM-mallin tavoitteita
Lainsäädäntö	Tulkinta ja rajat vaihtelevat, GDPR:n haasteet. Virossa tulkinta mahdollistaa ennustemallien käytön.	Tiukasti rajattua, ei selvää lakiperustetta tiedon vaihtamiselle	Tarvitaan juridista rajausta ns. suuremmalta julkiselta taholta
Organisointi	Keskitettyä Virossa, hajautetumpaa Alankomaat ja Tanska	Monialaista ja siiloutunutta, yksittäisiä kokeiluja, ei yhtä orgaisoivaa tahoa	Tarvitaan organisoiva taho, joka yhdistää toimijat

Eettiset näkökulmat	Ongelmallisia käyttöönottoja, joissa mm. profilointia ja GDPR:n virheellistä tulkintaa asiakkaan suostumukseen liittyen	Arkaluontoisten tietojen ja profiloinnin osalta ollaan tarkkoja	Läpinäkyvyys ja mallien selitettävyys keskeisiä huomioitavia seikkoja suunnittelussa ja toteutuksessa
---------------------	---	---	---

6.3.1 Toimijoiden rooli ja datanjakamisen kulttuuri

Alankomaissa viranomaiset jakavat dataa aktiivisesti, mikä tehostaa palveluiden koordinoitua, mutta samalla sääntely korostaa vastuullista ja läpinäkyvää tiedonhallintaa. Algoritmeja sekä sääntöksiä koordinoivat ja valvovat erikseen nimetyt viranomaiset. Alankomaissa käytössä oleva algoritmikehys antaa sääntöviivat ja ohjeistukset algoritmien käyttöönotolle. Sääntely painottaa vastuullista ja läpinäkyvää tiedonhallintaa. Esimerkiksi julkinen alusta algoritmirekisteri kokoaa viranomaisten käyttämät algoritmit yhteen, mikä lisää avoimuutta ja luottamusta. Rekisteri ei vielä kata kaikkia sektoreita, ja valtion tuottamia algoritmeja on vähän, mutta suunta on selvä, kohti avoimempaa hallintoa.

Tanskassa digitalisaatio on laajasti hallittu ja valtakunnallisesti koordinoitu kokonaisuus, jossa julkisen ja yksityisen sektorin yhteistyö on aktiivista. Epäonnistuneidenkin kokeilujen jälkeen digitalisaation edistämiseksi on tehty ja tehdään lainsäädännöllisiä toimenpiteitä. Tietosuojaviranomaisilla on merkittävä rooli, ja datan jakamisessa korostuu henkilötietojen minimointi ja eettinen vastuu. Digitaalinen suostumus ja sille rakennettu julkishallinnonkin käytössä oleva käyttöliittymä voi helpottaa tiedonjakamista jatkossa, roolit ja työnjako ovat kuitenkin vielä epäselviä.

Virossa X-Road on keskeinen osa julkisen hallinnon ja yksityissektorin välistä tiedonvaihtoa. Se toimii tietoturvallisena, standardoituna väylänä, jonka avulla eri organisaatiot voivat jakaa tietoa hajautetusti ja tehokkaasti. Virossa X-Road toimii kansallisena luottamusverkkona, johon kaikki julkishallinnon toimijat ja monet yksityiset palveluntarjoajat on liitetty. Viron lainsäädäntö velvoittaa viranomaisia käyttämään X-

Roadia, mikä on mahdollistanut laajan integraation ja lähes kaikki viranomaispalvelut ovat sen piirissä. Tiedon jakaminen on arkipäivää, ei poikkeus tai kokeilu.

Suomessa Datan jakamisen kulttuuri on varovainen, ja käytännössä tiedot kerätään ja yhdistellään harkiten, tiedonvaihto perustuu pitkälti organisaatiokohtaisiin ratkaisuihin. Vaikka Suomessa käytössä oleva Palveluväylä perustuu samaan X-Road-teknologiaan kuin Virossa, sen käyttö ja velvoittavuus ovat huomattavasti rajallisempia. Datan jakamista jarruttavat hajanaiset rakenteet sekä se, että toimijat eivät tunne riittävästi toistensa tietosisältöjä tai teknisiä muotoja. Vaikka esimerkiksi Kela, Väestötietojärjestelmä ja osa kunnista käyttävät Palveluväylää, sen käyttö ei vastaa Viron mallia. Suomessa Palveluväylä on monin paikoin jäänyt rajapintaratkaisuksi tai käytettäväksi vain tietyssä välttämättömässä tiedonsiirrossa, eikä se ole muodostunut kaiken kattavaksi infrastruktuuriksi. Suomessa ei ole myöskään selvää valvovaa eikä koordinoivaa tahoa julkishallinnon monitoimijaisille alustoille tai algoritmeille.

6.3.2 Teknologiset ja toiminnalliset edellytykset

Alankomaissa digitalisaatio ja data-analytiikka ovat vahvasti integroituneet julkisiin palveluihin. Alankomaissa ennustemallit ovat osa julkisia palveluita. Ennustemallit hyödyntävät monipuolisia taustamuuttujia ja ne on pyritty rakentamaan avoimesti, jolloin mallien toiminnasta pyritään antamaan selkeää tietoa käyttäjille. Teknologisesti Alankomaiden järjestelmät tukevat joustavaa tiedonvaihtoa, mikä edesauttaa ennustemallien käytön laajentamista sosiaaliturvan palveluissa. Teknologia mahdollistaa laajan datan hyödyntämisen, mutta SyRI-tapaus muistutti, kuinka vakavia seurauksia voi olla, jos algoritmien käyttö on läpinäkymätöntä ja syrjivää.

Tanska on ottanut tekoälyn käyttöön julkisella sektorilla aktiivisesti. Tätä on tukenut selkeä kansallinen strategia sekä erityisen ohjaavan toimielimen perustaminen. Tanskassa panostetaan julkisen IT-infrastruktuurin jatkuvaan kehittämiseen, mutta ASTA-työkalu osoitti kuitenkin, miten tärkeää on säilyttää luottamus ja läpinäkyvyys ennustemallien käytössä, sillä algoritmien toiminnan ymmärtämättömyys ja tietosuojaan liittyvät epäselvyydet johtivat työkalun käytön lopettamiseen. Teknologisesti Tanska panostaa julkisen IT-infrastruktuurin jatkuvaan uudistamiseen, mutta kokemukset ASTAsta muistuttavat varovaisuuden tarpeesta ennustemallien käyttöönotossa. Palveluohjauksen tukena ennustemallit oli ajateltu resurssien kohdentamiseen, mutta lopulta työkalun käyttö

jäi puutteellisen hyväksynnän takia rajalliseksi. ASTA on jäänyt esimerkiksi siitä, miten julkisen sektorin tekoälytyökalut voivat epäonnistua, jos läpinäkyvyys, käyttäjien koulutus ja oikeudelliset reunaehdot eivät ole riittävästi huomioitu kehitystyössä. Vaikka algoritmin tavoite oli tukea varhaista puuttumista ja ennaltaehkäisyä työmarkkinoilla, sen käyttöönotto osoitti, kuinka vaikeaa on yhdistää koneoppiminen eettisesti ja käytännöllisesti julkiseen palveluun.

Virossa digitaalinen palveluympäristö on poikkeuksellisen kehittynyt muihin maihin verrattuna ja lähes kaikki kansalaisasiointi on siirtynyt verkkoon. X-Road-infrastruktuuri mahdollistaa turvallisen ja sujuvan datanvaihdon viranomaisten välillä ilman tarpeetonta tiedon monistamista. Viro on onnistunut yhdistämään teknologisen edistyksen, datan laajan käytön ja yksityisyydensuojan siten, että kansalaisten tiedot ovat suojassa, mutta viranomaiset voivat toimia tehokkaasti. Turvallinen ja luotettava tiedonvaihto eri toimijoiden välillä mahdollistetaan avoimen lähdekoodin KSI-lohkoketjuteknologialla, joka takaa tietojen eheyden ja jäljitettävyyden. Virossa ennustemallit, kuten OTT, nojaavat tehokkaaseen tiedonvaihtoon ja vakiintuneeseen teknologiseen infrastruktuuriin. Mallien käyttö on osa viranomaisten arkea, ei erillinen projekti. Teknologia tukee ketterää viranomaistyötä ja nopeaa palveluohjausta.

Teknologisesti Suomessa rakennetaan kehittyneitä koneoppimismalleja, mutta tiedonvaihto ja ymmärrys muiden datasta on vähäistä. Toiminta on eriytynyttä, ja ennustemallien kehitys lähtee usein organisaatiokohtaisista tarpeista. Ennustemallit voivat tukea palveluohjausta tarjoamalla työnhakijoiden riskiprofiileja, joiden avulla resursseja voidaan kohdentaa tehokkaammin, mutta päätökset tehdään aina ammattilaisten arvioinnin pohjalta. Ennustemallien kehittäminen keskittyy vielä tällä hetkellä ainoastaan organisaation sisäisen tai jo käytettävissä olevan datan käyttöön ja nykytiedon pohjalta voidaan arvioida, että datan onnistunut yhdistäminen eri toimijoiden välillä parantaisi edelleen ennustemallien ennustetta. Tältä osin tekniset ja hallinnolliset esteet ovat yhä ratkaisematta. Teknologista osaamista siis löytyy, mutta toiminnalliset edellytykset kehittämiselle ja käyttöönotolle puuttuvat.

6.3.3 Lakien soveltaminen ja tulkinta

Alankomaissa Data Actin toimeenpano on selkeästi organisoitu, kuluttajaviranomainen ja tietosuojaviranomainen on nimetty vastuullisiksi viranomaisiksi. GDPR on asettanut

rajoitteita ja sen noudattamatta jättäminen tai väärä tulkinta on johtanut huomautuksiin sekä joidenkin käyttötapausten lakkauttamiseen. Alankomaiden SyRI-tapauksen kuvaus on varoittava esimerkki siitä, miten epäonnistunut algoritmien ja rekisteritietojen käyttö voi johtaa vakaviin oikeudellisiin ja eettisiin ongelmiin, erityisesti, jos järjestelmän toiminta ei ole läpinäkyvää tai se kohdistuu epätasaisesti eri väestöryhmiin. Järjestelmä yhdisteli laajasti kansalaisten henkilötietoja useista viranomaisrekistereistä, esimerkiksi asumisesta, tuloista, sakoista ja sosiaalietuuksista, tunnistaakseen henkilöitä, jotka poikkeavat tavanomaisista profiileista ja jotka siksi pitäisi tutkia tarkemmin. Suurimmiksi ongelmiksi muodostui läpinäkyttömyys, eli algoritmi oli salainen, eikä kansalaisilla ollut pääsyä tietoihin siitä, miksi heidät oli luokiteltu riskitapauksiksi, sekä syrjivät käytännöt, kun järjestelmää käytettiin erityisesti matalatuloisilla alueilla, mikä synnytti vaikutelman valikoivasta valvonnasta ja sosiaalisesta leimauttamisesta. Kansalaisilla ei ollut tehokasta keinoa puolustautua järjestelmän tekemiä luokituksia vastaan.

Automaattista päätöksentekoa koskeva sääntely on jo pitkään Tanskassa pyritty sisällyttämään sektorikohtaisiin lakeihin, minkä on tarkoitus mahdollistaa tehokkuushyödyt viranomaisten työssä. Tanskassa GDPR:n tulkinta on osin rajoittanut ennustemallien laajaa hyödyntämistä. ASTA:n valossa GDPR:ää tulkitaan tiukasti julkisessa hallinnossa, erityisesti kun kyse on automatisoidusta arvioinnista tai päätöksenteon tukemisesta. Suostumusta ei voida pitää GDPR:n mukaisena laillisena perusteena tietojen käsittelylle, koska kansalaisen suostumusta ei voi pitää ”vapaaehtoisena”, jos suostumuksen antaminen on käytännössä ehto palvelujen saamiselle. Tämä asettaa rajoitteita varsinkin automatisoiduille ratkaisuille julkisissa palveluissa. Tanskan tämänhetkinen työllisyysuudistus ja digitalisaation toimenpiteet voivat kuitenkin antaa hyvät lähtökohdat työllisyyspyrkimyksien edistämiseen tekoälyn ja koneoppimisen keinoin.

Virossa GDPR-tulkinta on tarkka mutta samalla mahdollistava. Digitaalinen infrastruktuuri on suunniteltu siten, että tietosuojaperiaatteet, kuten minimointi, on rakennettu järjestelmän ytimeen. Tämä on mahdollistanut vahvan EU-lainsäädännön noudattamisen ilman, että tehokas datan käyttö vaarantuu. Lainsäädäntö tukee digitaalista toimintaympäristöä ja varmistaa tietojen käytön läpinäkyvyyden. Virossa yksityisten palveluntarjoajien ja julkisen sektorin välinen yhteistyö on ollut vähäisempää, koska tiukat tietosuoja- ja tietoturva vaatimukset rajaavat pääsyä julkisiin rekistereihin. Lainsäädäntö edellyttää, että tietoja voivat käyttää vain ne tahot, joilla on siihen laissa määritelty peruste. Palveluohjauksessa ennustemallit ovat keskeisiä työkaluja, jotka auttavat resurssien

kohdentamisessa ja nopeuttavat asiakkaiden palvelupolkua. Viron mallissa olisi paljon opittavaa, jotta Suomessa voitaisi päästä edes lähelle eri toimijoiden tiedonvälitystä.

Suomessa GDPR-tulkinta on hyvin varovaista, erityisesti kun kyse on datan jakamisesta toimijoiden välillä ja suostumuksen antamisesta. Toimijoiden roolissa korostuu julkisen sektorin vastuullisuus, ja tietojen käyttö perustuu tarkkaan lainsäädäntöön, joka huomioi yksityisyyden suojan sekä eettiset periaatteet. Ennustemallien kehityksessä käytetään vain oman organisaation dataa laillisten rajoitusten puitteissa. Tämä suojaa asiakkaan oikeuksia, mutta samalla se voi rajoittaa teknologian potentiaalia, jos datan käyttö ja vaihto jäävät hajanaisiksi. Vaikka toisiolaki säätelee sosiaali- ja terveystietojen käyttöä muuhun kuin alkuperäiseen käyttötarkoitukseen, siitä ei ole olemassa selkeitä ennakkotapauksia, eikä laki sellaisenaan tarjoa suoraa ratkaisua tiedon jakamiseen. TYM-lain toivottiin tuovan selkeyttä tiedonjakamiseen viranomaisten välillä, erityisesti TYM-yhteistoimintamallin osalta. Lainsäädännön tulkinta siitä, mitä tarkoitetaan henkilön työllistymiseen vaikuttavilla ja palvelujen tarjoamisen kannalta välttämättömillä tiedoilla, on edelleen epäselvä, eikä anna varmaa vastausta ja keinoa toimijoidenväliseen tiedonjakoon. Yhteistyötä suunniteltaessa on otettava huomioon myös sellainen TYM-lakiin liittymätön yleislainsäädäntö, joilla voi olla vaikutusta käytännön toteutukseen. Tämän opinnäytetyön tarkastelu ei ole tyhjentävä. Tekoälyn keskeistä lainsäädäntöä Suomen näkökulmasta on käsitelty luvussa 4.5.

7 Vertailun tulokset

Vertailun jälkeen voidaan todeta, että kaikissa kolmessa maassa ennustemalleja käytetään sosiaaliturvan palveluohjauksen tukena, mutta eri maiden välillä on merkittäviä eroja digitaalisen infrastruktuurin, tiedonvaihdon ja lainsäädännöllisten tulkintojen osalta. Tarkasteltujen maiden esimerkit korostavat, että ennakoiva sosiaali- ja terveydenhuolto edellyttää moniammatillista yhteistyötä, vahvaa strategista ohjausta sekä teknologisten ratkaisujen kehittämistä ja hyödyntämistä. Näiden kokemusten pohjalta Suomella on mahdollisuus tunnistaa toimivia käytäntöjä ja soveltaa niitä omiin olosuhteisiinsa tukemaan vaikuttavampaa, kestävään pohjaan rakentuvaa ja ihmislähtöistä sote-järjestelmää. Eri maiden alustat ja julkisten palvelujen rakenne eroavat toisistaan lähtökohtaisesti jo ennen dataekosysteemien rakentamista. Kuitenkin muiden maiden esimerkit tuovat ilmi tärkeitä huomioita niin onnistumisten kuin epäonnistumistenkin kannalta. Jo suunnittelun

varhaisessa vaiheessa on oleellista varautua riskeihin, sekä ottaa oppia onnistuneesta datatalouden yhteistyöstä ja ennustemallien toteutuksista. Onnistunut digitalisaatio parantaa prosessien tehokkuutta, mahdollistaa uusien liiketoimintamallien syntyminen ja edistää innovaatioita. Julkishallinnon tasolla digitalisaatio näkyy onnistuneissa esimerkkitapauksissa palvelujen parempana saatavuutena ja resurssien säästönä. Alankomaiden Work Profilerin myötä on todettu, että arvio on kokonaisuutena myös tarkempi kuin asiantuntijan työ yksinään. Hallinnollisen datan jakaminen eri lähteistä helpottaa tilanteissa, joissa asiakkailta on monimutkaisia palvelutarpeita eri toimijoiden välisissä kokonaisuuksissa.

Ennustemallien kehittämistä ja käyttöönottoa sosiaaliturvan palveluohjauksessa hidastavat edelleen avoimet juridiset ja eettiset kysymykset. Kirjallisuuskatsauksessa korostuu, että luottamus ja läpinäkyvyys ovat keskeisiä edellytyksiä tekoälyn vastuulliselle hyödyntämiselle. Maiden väliset erot ennustemallien käytössä heijastavat erilaisia hallinnollisia, oikeudellisia ja kulttuurisia lähtökohtia. Suomessa ja Tanskassa painotetaan eettisyyttä ja lakien tulkinta on tiukkaa, mikä näkyy varovaisuutena sekä tietosuojaan ja suostumuksen vahvana painottamisena. Alankomaissa lähestymistapa on joustavampi, ja ennustemalleja sekä niihin liittyviä algoritmeja jaetaan avoimesti, mikä mahdollistaa laajemman tiedon hyödyntämisen. Viro puolestaan erottuu keskitetysti hallitulla ja teknologisesti pitkälle kehitetyllä infrastruktuurillaan, jossa datan hallinta on integroitu osaksi julkisia palveluja turvallisella tavalla. Näistä kokemuksista voidaan päätellä, että ennustemallien hyödyntämisen edellytyksenä on yhtä aikaa toimiva teknologia, selkeä lainsäädännöllinen kehys sekä luottamuksen rakentaminen niin ammattilaisten kuin kansalaistenkin suuntaan.

Digitaalisen infrastruktuurin ratkaisut vaikuttavat merkittävästi siihen, missä määrin ennustemalleja voidaan käytännössä hyödyntää. Alankomaissa julkisten palvelujen ja puolijulkisten tai yksityisten palvelujen välillä on selkeät rajat digitaalisten valtion infrastruktuurien käytössä. Virossa sen sijaan on rakennettu yhtenäinen ratkaisu, X-Road, jota hyödynnetään tehokkaasti sekä julkisissa että yksityisissä palveluissa. Suomessa ei ole vastaavaa keskitettyä digitaalista infrastruktuuria, joka mahdollistaisi yhtä laajan tiedonjakamisen eri toimijoiden välillä. Tiedonvaihto perustuu hajanaisiin järjestelmiin ja vanhoihin rajapintoihin, joita ei ole suunniteltu ennustemallien tarpeisiin. Tämä rajoittaa mallien käyttömahdollisuuksia ja vaikuttavuutta. Tästä syystä on aiheellista arvioida, onko nykyinen infrastruktuuri riittävän yhtenevä ja millaisia kehittämistoimia tarvitaan yhteisen

tiedonjaon ja sitä kautta palveluohjauksen tueksi. Jää nähtäväksi voisiko esimerkiksi SOTE-tekoälyn ekosysteemiverkosto kehittää yhteisiä alustoja tai auttaa yhdistämään rakenteita.

Kansainväliset esimerkit havainnollistavat erilaisia lähestymistapoja tekoälyn käyttöönottoon julkishallinnossa ja sosiaaliturvassa. Alankomaat ja Viro sijoittuvat Global Index on Responsible AI -listan kärkikymmenikköön, mikä kuvastaa niiden pitkälle kehittyneitä digitaalista hallintoa ja vastuullista tekoälypolitiikkaa. Alankomaat painottaa avoimuutta algoritmien käytössä, kun taas Viro on rakentanut digitaalisen palveluarkkitehtuurinsa tavalla, joka minimoi tarpeettoman tiedon käsittelyn ja kopioimisen. Viron hallinto korostaa lisäksi kokeilun ja käytännön oppimisen merkitystä, mikä eroaa monien muiden maiden varovaisemmasta lähestymistavasta. Erityisesti Suomessa suhtaudutaan varovaisesti riskeihin ja dataekosysteemikokeilut ovat pieniä yleensä yksityisten toimijoiden välisiä. Tanskassa ja Alankomaissa on kuitenkin havaittu ongelmia tilanteissa, joissa mallien logiikkaa ei ole avattu riittävän läpinäkyvästi käyttäjille tai ne eivät täyttäneet tietosuoja-asetuksen vaatimuksia, mikä on heikentänyt luottamusta teknologiaan.

Ennustemallien ja muiden tekoälyratkaisujen käyttöönotto palveluohjauksessa edellyttää syvällistä juridista analyysiä erityisesti tietosuojalainsäädännön ja tulevan tekoälysääntelyn näkökulmasta. EU:n yleinen tietosuoja-asetus (GDPR) edellyttää, että automatisoitua päätöksentekoa ja profilointia koskevat riskit arvioidaan huolellisesti esimerkiksi vaikutustenarvioinnin avulla. Lisäksi on varmistettava, että rekisteröidyllä säilyy oikeus tulla kuulluksi ja saada ymmärrettävä selitys päätöksenteon perusteista. Vaikka automatisoitua päätöksentekoa ei vielä ole tarkastelluissa maissa käytössä, järjestelmien oikeudellinen arviointi tulee tehdä huolellisesti jo suunnitteluvaiheessa. Tämä koskee myös uutta EU:n tekoälysäädöstä (AI Act), jonka mukaan korkean riskin tekoälyjärjestelmien, kuten työllisyyspalveluissa käytettävien ennustemallien, on täytettävä muun muassa läpinäkyvyyttä, valvottavuutta ja tietoturvaa koskevat vaatimukset. Näin ollen ennustemallien käyttöönotto edellyttää sekä tietosuojan että tekoälysääntelyn näkökulmasta dokumentoitua, läpinäkyvää ja oikeudellisesti kestäväää lähestymistapaa.

EU:n yleisen tietosuoja-asetuksen (GDPR) tulkinta vaikuttaa merkittävästi ennustemallien kehittämiseen ja käyttöönottoon. Viron tulkinta ja soveltaminen GDPR:stä on erityisen mielenkiintoinen siksi, että maa on samanaikaisesti sekä digitaalisen hallinnon edelläkävijä

että erittäin vahvasti EU-oikeuteen sitoutunut valtio. Suomessa korostetaan erityisesti tietosuojaa ja rekisteröidyn suostumuksen merkitystä, ja GDPR:n sekä muiden lakien tulkinta on varovaista. Vaikutustenarviointi (DPIA) on monesti osoittautunut hidastavaksi tekijäksi, sillä prosessi on aikaa vievä ja alkaa aina alusta. Olisi perusteltua, että kartoittavalle ja kokeilevalle vaiheelle olisi olemassa kevyempi arviointimalli, kun taas tuotantoon tähtääville ratkaisuille tehtäisiin perusteellinen ja huolellinen vaikutustenarviointi. Tietosuojan näkökulmasta ainoa varma keino välttää henkilötietojen käsittely on tällä hetkellä anonymisointi, sillä pelkkä pseudonymisointi ei riitä estämään yksilön epäsuoraa tunnistamista. Palveluohjaukseen tarkoitettussa ennustemallissa on kuitenkin tunnistettava palvelutarpeessa oleva henkilö. Lainsäädännöllisesti tarvitaan selkeitä linjauksia myös siitä, mihin TYM-laki Suomessa oikeuttaa. Jos TYM-laissa säädetään, että voidaan käyttää sellaisia terveydentilaa ja työ- ja toimintakykyä koskevia tietoja, jotka ovat välttämättömiä palvelujen tarjoamiseksi, tarvitaan selkeitä linjauksia mitä nämä välttämättömät tiedot ovat. Lisäksi rekistereihin saa tallettaa tietoja siltä osin, kun ne ovat tarpeellisia palvelujen tarjoamiseksi. Kun puhutaan palvelujen tarjoamisesta, jää epäselväksi voidaanko ennustemallien kehittäminen katsoa osaksi palvelukokonaisuutta.

Maiden välisessä vertailussa tunnistetut epäonnistumiset korostavat sitä, että ilman riittävää eettistä ja lainsäädännöllistä valmistautumista ennustemallien käyttö voi johtaa ongelmiin. Suurissa massoissa pienikin virhemarginaali voi johtaa lukuisiin virheellisesti tunnistettuihin henkilöihin tai epäsuotuisaan profilointiin. Eettiset näkökulmat on otettava huomioon myös siinä, kuinka paljon painoarvoa tuloksille annetaan ja miten tuloksia hyödynnetään. Ennustemallin rajaus ei saa jättää palvelutarpeessa olevia asiakkaita ulos palvelujen piiristä. TYM-yhteistoimintamalliin ei voi tuoda algoritmipohjaista tukea ilman, että mallien toiminta on täysin läpinäkyvää ja että niitä käytetään vain päätöksenteon tukena, ei automaattisesti päätöksenteossa.

Tanskassa ja Alankomaissa ennustemalleja on kehitetty ja kokeiltu palveluohjauksen tueksi. Molemmissa maissa mallit tuottavat taustatietoa, jotka auttavat ammattilaisia arvioimaan asiakkaiden palvelutarpeita ja kohdentamaan resursseja tarkoituksenmukaisesti. Ennustemallit eivät kuitenkaan ohjaa asiakkaita automaattisesti palveluihin tai johda automatisoituihin päätöksiin. Suomessa TYM-malli perustuu tällä hetkellä täysin ihmistyöhön ja monialaiseen keskusteluun, eikä ennustemalleja hyödynnetä vielä asiakasohjauksen tukena. Ennustemallit voisivat toimia TYM-mallin sisällä taustatietona, esimerkiksi tunnistamassa varhaisen tuen tarpeessa olevia henkilöitä, mutta

niiden käyttö vaatisi vahvaa selitettävyyttä ja oikeudellista perustaa, jotta ihmiskeskeisyys ja asiakkaan oikeusturva säilyvät. Virossa ennustemallien käyttö perustuu laajaan viranomaisten väliseen tiedonvaihtoon. Virossa ennustemalli tuottaa yhteisrekisteristä taustatietoa, jota voidaan hyödyntää monialaisen tuen näkökulmasta. Viron esimerkistä voidaan oppia, että vahvasti integroitu tietoinfrastruktuuri on ennustemallien operatiivisen käytön edellytys, ilman sitä algoritmien soveltaminen jää kokeiluasteelle tai yksittäisiksi projekteiksi. Kelan ja Helsingin kaupungin työkykykokeilu osoittaa, että mallien kyky havaita riskejä on teknisesti hyvä, mutta käytännön vaikuttavuus jää vähäiseksi, koska ennuste tulee liian myöhään eikä sen ole mahdollista ohjata palvelupolkua monen toimijan näkökulmasta. Tiedon hajanaisuus rajoittaa osaltaan yhteisten ennustemallien kehittämistä. Tietoa on eri muodoissa, rakenteet tiedon vaihtoon ovat hajanaiset ja käytössä on vanhoja rajapintoja, joita ei ole tarkoitettu ennustemallien käyttövaatimuksiin. Ennustemallien hyödyntäminen TYM-palvelussa olisi tehokkaampaa, jos tietojen jakamista TE-toimiston, Kelan ja kuntien välillä voisi helpottaa turvallisesti, yhteisen alustan laajemman käytön sekä lainsäädännön selkiyttämisen avulla.

Lukuisat erot tekevät haastavaksi kopioida suoraan toisen maan parhaita käytäntöjä, sillä eri maissa toimivat instituutiot ja tiedonvaihtojärjestelmät poikkeavat toisistaan. Silti jokaiselta maalta löytyy arvokasta opittavaa. On tärkeää, että luotettavat tahot pystyvät vastaamaan odotuksiin niin, että osaaminen on ajan tasalla sekä palvelut toimivat sujuvasti. Julkisen sektorin vastuullisuus on tässä myös keskeisessä roolissa. Tarvitsemme muutosta siihen, miten hallinto toimii, siirtymistä pois sisäänpäin kääntyneistä ja hajanaisista organisaatorakenteista kohti tiiviimpää yhteistyötä. Vasta silloin voimme tarjota kansalaisille ja yrityksille sujuvia, kokonaisvaltaisia palvelukokemuksia, joissa eri toimijat toimivat saumattomasti yhdessä. Digitaalisten teknologioiden hyödyntäminen maksimaalisesti vaatii myös sopivia tiedonvaihtoinfrastruktuureja yhteentoimivuuden ja digitaalisten palvelujen turvaamiseksi julkisella sektorilla. Tämä edellyttää yhteisiä standardeja, tiedon jakamista, osaavaa henkilöstöä ja organisaatioiden sisäistä osaamista, sekä johtamista ja rakenteita, joita ei vielä ole olemassa. Useimmat merkittävät tiedonvaihdon käyttötapaukset kytkeytyvät tavalla tai toisella henkilötietoihin. Julkishallinnossa näiden tietojen käsittely yli hallintorajojen on monimutkaista, mutta välttämätöntä, jotta palvelut toimivat kokonaisuutena. Turvallisen ja lainmukaisen tiedonkäsittelyn tueksi tarvitaan kansallinen hallintamalli, joka tarjoaa selkeät pelisäännöt ja vahvistaa toimijoiden kykyä hyödyntää dataa myös palvelujen kehittämisessä ja tuotteistamisessa. Selvä hallintamalli tukisi läpinäkyvyyttä ja vahvistaisi käyttäjien ja

kansalaisten luottamusta. Onkin tärkeää käydä avointa keskustelua ja hakea kansalaisten hyväksyntää sille, että tietojen käyttö ja yhdistely on sallittua. Samalla on rakennettava luottamus siihen, että tietoja hyödynnetään vain siihen tarkoitukseen, johon ne on kerätty.

Vaikka maat ovat ilmoittaneet kansallisesta tekoälystrategiastaan, ei käytännön jalkauttaminen ole niin yksinkertaista. Yksittäisten toimijoiden vaikutusmahdollisuudet ovat pienet. Tarvitaan suurempi koordinoiva taho, erityisesti julkishallinnon ja sosiaaliturvan toimijoiden avuksi. Toisilain ennakkotapauksia ei ole, eikä monella ole halua juridiseksi koekaniiniksi ryhtyä, eivätkä julkiset resurssit ja rahoitus edes aina mahdollista pilotointivaihetta, vaikka kokeilujen pohjalta olisi saatu lupaavia tuloksia. Suomessa olisi syytä arvioida tarkemmin, onko viranomaissektorilla tarvetta erilliselle toimielimelle, joka ohjaisi ja valvoisi tekoälyn käyttöä. Kansainväliset esimerkit voivat tukea tällaisen rakenteen perustamista, erityisesti jos toimielin pystyisi keskitetyksi tarjoamaan juridista ja eettistä ohjeistusta tekoälyyn liittyvissä kysymyksissä. Asian jatkoselvittäminen olisi kuitenkin tärkeää, erityisesti toimivaltakysymysten ja vastuunjaon osalta.

7.1 Kehittämisessä huomioitavaa

Tähän lukuun on pyritty teoriapohjan ja vertailun perusteella tekemään tiivistelmä siitä, mitä ennustemallin monitoimijaisessa kehittämisessä on huomioitava jo ennen itse mallin toteuttamista. Sosiaali- ja terveysalan data on monimuotoista, ja se voi sisältää numeerisia arvoja, koodattua tietoa, vapaamuotoista tekstiä sekä kuvia. Eri tietotyypit vaativat omia käsittelymenetelmiään, ja mallinnuksen onnistuminen edellyttää teknisen osaamisen lisäksi ymmärrystä datan merkityksestä ja käyttökontekstista, mikä puolestaan vaatii organisaation sisäisen datanhallinnan toimivuutta.

Henkilöstön osaaminen on ratkaisevaa. Asiantuntijan tulee tietää, mistä tiedot ovat peräisin, mihin niitä saa käyttää ja mitä vastuita käyttöön liittyy. Tämä edellyttää koulutusta, joka kattaa sekä algoritmien peruseräatteen että datan oikeudelliset ja sisällölliset ulottuvuudet. Ilman riittävää ymmärrystä on vaarana, että mallien tuottamia tuloksia tulkitaan virheellisesti, mikä voi johtaa virheellisiin johtopäätöksiin tai epäoikeudenmukaiseen kohteluun. Ennustemallien läpinäkyvyys ja selitettävyys eivät yksin riitä, jos henkilöstöllä ei ole valmiuksia ymmärtää niiden toimintaa. Dokumentointi, esimerkiksi tekoälyn koulutuksessa käytetyn aineiston, algoritmin ja testitulosten huolellinen

kirjaaminen ja säilytys, on keskeistä oikeudellisesti kestäväen suunnittelun kannalta. Suuririskisiä järjestelmiä koskeva sääntely asettaa omat velvoitteensa, mutta kehitystä voitaisiin tukea myös sääntelyhiekkalaatikoiden avulla. Ne tarjoavat mahdollisuuden testata uusia ratkaisuja hallitusti ja arvioida niiden oikeudellisia vaikutuksia ennen laajempaa käyttöönottoa.

Kun ennustemallia hyödynnetään osana palveluprosessia, asiakkaan tiedottaminen on olennaista. Hänelle on kerrottava ymmärrettävästi, miten hänen tietojaan käytetään, mihin tarkoitukseen ja vaikuttaako tämä esimerkiksi kohdennettaviin palveluihin. Avoin tiedottaminen lisää luottamusta ja mahdollistaa asiakkaan oikeuksien käytön. Suomessa vahva luottamus digitaalisiin ratkaisuihin antaa hyvän pohjan, mutta hyväksyntää tekoälyn käytölle on rakennettava myös julkisen keskustelun kautta. Taulukossa (Taulukko 2) on koottuna tiivistettynä kehittämisen kannalta keskeiset huomiot.

Taulukko 2. Suunnittelussa huomioitavaa

Osa-alue	Kuvaus	Huomiot
Datan hallinta kuntoon	Eri tietotyypit vaativat omia käsittelymenetelmiä.	Tarvitaan teknistä osaamista ja ymmärrystä datan merkityksestä ja kontekstista.
Henkilöstön kouluttaminen	Asiantuntijoiden on tiedettävä datan alkuperä, käyttötarkoitus ja vastuut.	Koulutuksen tulee kattaa algoritmien peruseräatteen sekä datan oikeudelliset ja sisällölliset ulottuvuudet. Henkilöstön luottamus ja ennustemallin ymmärrys onnistumisen keskiössä.

Dokumentointi	Koulutuksessa käytetyn aineiston lakisääteinen dokumentointi ja säilytys.	Selvitettävä myös mitä dokumentointia vaaditaan jo kokeiluvaiheessa. Helpotusta mahdollisilla sääntelyhiekkalaatikoilla.
Sääntely	Suuririskinen järjestelmä, joten tietyt veloitteet ja vaatimukset.	Riskienhallinta, datanhallinta, dokumentointi ja valvonta. Viranomaisen hyväksyntä käyttöönottoon.
Tiedottaminen	Asiakkaalle kerrottava miten hänen tietojään käytetään, mihin tarkoitukseen ja miten se vaikuttaa palveluihin.	Lisää luottamusta, mahdollistaa asiakkaan oikeuden pyytää lisätietoa tietojenkäytön perusteista.
Julkinen keskustelu	Saavutettava kansalaisten hyväksyntä tietojen käyttöön	Luottamus palveluihin

7.2 Mahdolliset riskitekijät

Tähän lukuun on pyritty teoriapohjan ja vertailun perusteella tekemään tiivistelmä siitä, mitä mahdollisia riskitekijöitä ennustemallin monitoimijaisessa kehittämisessä on otettava huomioon. Tiivistelmä ei ole tyhjentävä, mutta se pyrkii välttämään maiden vertailussa esiintulleiden epäonnistuneiden käytötapausten kaltaiset eettiset ja oikeudelliset ongelmat. Ennustemallin näkökulmasta sosiaali- ja terveydenhuollossa on useita datankäyttöön liittyviä haasteita. Eri rekistereissä olevan yksilötason tiedon yhteiskäyttö edellyttää selkeää lainsäädännöllistä perustaa, mutta nykyinen toisiolaki tai TYM-laki eivät anna yksiselitteistä oikeudellista perustaa tiedonvaihtoon. TYM-laki itsessään ei säädi esimerkiksi rekistereiden yhdistämisestä. Tarve yhteiselle koordinoivalle taholle on ilmeinen, jotta

voitaisiin määritellä, mitkä tiedot ovat välttämättömiä yhteistoimintamallin toteuttamiseksi. Onko ennustemalli osa palvelua, joka kuuluu yhteistoimintamalliin ja jonka perusteella voidaan jakaa dataa eri toimijoiden välillä? Tämän lisäksi ongelmana on, ettei eri toimijoilla ole tietoa toistensa hallussa olevasta datasta, mikä vaikeuttaa potentiaalisen tunnistamista ja hidastaa innovointia. Myös kokeilujen läpivieminen on usein hidasta, sillä tietotarpeet tarkentuvat vasta käytännön testien aikana ja jokainen vaihe vaatii erillisen vaikutustenarvioinnin.

Teknisistä näkökulmista merkittävä riski liittyy tekoälyn niin sanottuun musta laatikko - ilmiöön. Neuroverkkoihin ja syväoppimiseen perustuvien mallien päätöksentekoa on usein mahdotonta selittää tai perustella läpinäkyvästi, mikä ei vastaa lainsäädännön dokumentointivaatimuksia. Tämä herättää kysymyksen, kuinka ennustemallien tuottamaa tietoa voidaan verrata perinteisin menetelmin tuotettuun tietoon. Jos kuitenkin voidaan luotettavasti osoittaa, että tekoäly antaa yhtä luotettavia tuloksia kuin ihminen, olisi tämä syytä huomioida sääntelyä kehitettäessä. Samalla tarvittaisiin yhteisiä viranomaisohjeita ja koordinaatiota, jotta tulkintaerot esimerkiksi tietosuojan soveltamisessa eivät estäisi mallien hyödyntämistä. Lisäksi riskejä liittyy eettisiin ja käytännöllisiin näkökulmiin. Ennustemallit voivat profiloida asiakkaita epäedullisesti tai kohdentaa palveluita liian suppeasti, jolloin osa todellisista avuntarvitsijoista jää ulkopuolelle. Asiakkaan suostumusta ei myöskään voida pitää riittävänä perusteena tietojen käyttöön, sillä palveluiden saaminen ei voi olla riippuvaista suostumuksen antamisesta. Resurssien riittävyys muodostaa oman ongelmansa, jos malli nostaa esiin aiempaa enemmän palvelun tarpeessa olevia, on vaarana, ettei tarvittavia palveluita riitä kaikille. Ennustemallien rajauksessa tulisi siksi huomioida taloussuhdanteiden kaltaiset tekijät. Näiden kysymysten ratkaiseminen on välttämätöntä, jotta tekoälyä voidaan hyödyntää oikeudenmukaisesti ja kestäväällä tavalla sosiaali- ja terveyspalveluissa. Taulukossa (Taulukko 3) on koottuna tiivistetysti mahdolliset riskitekijät sekä ratkaisut.

Taulukko 3. Mahdolliset riskitekijät

Riskikategoria	Kuvaus	Mahdollinen ratkaisu
----------------	--------	----------------------

Lainsäädännön haasteet	TYM-laki tai toisiolaki ei anna selvää perustetta tietojenkäytölle	Kansallinen yhteinen linjaus rekisterien yhdistämisestä, linjattava TYM-lain osalta käyttötarkoitus
Tietojen läpinäkyvyys	Ei tietoa muiden datasta	Yhteistyö ja mahdolliset metatietokatalogit
Kokeiluvaiheen esteet	Toteutus hidasta, DPIA raskas prosessi, tietotarpeet tarkentuvat myöhässä	Valvotut testiympäristöt kuten sääntelyhiekkalaatikot
Perusteltavuus	Ongelmana musta laatikko - ilmiö ja onko päätös luotettava	Explainable AI ja verrattavuus asiantuntijan tuottamaan tulokseen
Eettisyys	Riskinä profilointi tai palvelujen kohdentuminen väärin	Vaikutustenarviointi, suostumus ei saa olla palvelunsaannin edellytys
Mallin rajaus	Malli nostaa systemaattisesti kaikki palvelutarpeessa olevat, määrällisesti mahdollista kasvua	Ennakoitu resurssisuunnittelu, riittävä mallin kalibrointi

8 Johtopäätökset ja pohdinta

Tässä luvussa kootaan yhteen viitekehyksen ja kirjallisuuskatsauksen keskeiset havainnot, ja arvioidaan niiden perusteella vastauksia työn alussa esitettyihin tutkimuskysymyksiin:

1. Millaisia oikeudellisia, teknisiä ja organisatorisia edellytyksiä datan jakamiseen ja yhteentoimivuuteen liittyy julkishallinnon ennustemallien käytössä?
2. Kuinka tiedonjaon rajoitukset ja yhteisrekisterinpito vaikuttavat ennustemallien hyödyntämiseen TYM-palveluiden näkökulmasta, eli mitä dataa voidaan käyttää?
3. Miten ennustemallien käyttöä on toteutettu muissa EU-maissa työllisyyden tukemisessa ja työkyvyttömyyden ehkäisyssä, ja mitä oppeja niistä voidaan soveltaa Suomeen?
4. Miten varmistetaan ennustemallien eettisesti ja oikeudellisesti kestävä käyttö monitoimijaisessa yhteistyössä?

Vertailussa (luku 6.3) pyrittiin tarkastelemaan tutkimuskysymystä siltä osin, miten ennustemallien käyttöä on toteutettu muissa EU-maissa työllisyyden tukemisessa ja työkyvyttömyyden ehkäisyssä, ja mitä oppeja niistä voidaan soveltaa Suomeen. Vertailun tuloksissa (luku 7) havaittiin, että ennustemalleja käytetään EU-maissa sosiaaliturvan palveluohjauksen tukena, mutta maiden välillä on eroja digitaalisessa infrastruktuurissa, tiedonvaihdossa ja lainsäädännön tulkinnoissa. Suomessa voitaisiin hyödyntää erityisesti Viron yhtenäistä digitaalista infrastruktuuria, Alankomaiden avoimuutta algoritmien käytössä sekä muiden maiden kokemuksia moniammatillisesta yhteistyöstä ja luottamuksen rakentamisesta ennustemallien vastuullisessa käytössä.

Tekoälyasetusten vaikutusten arviointi on Suomessa yleisellä tasolla vielä kesken ja juridiset linjaukset ovat tulkinnanvaraisia. Tämä luo epävarmuutta erityisesti tilanteissa, joissa ennustemalleja kehitetään ja käytetään usean julkishallinnon toimijan yhteistyönä, ja joissa dataa joudutaan yhdistämään eri rekisterinpitäjien järjestelmistä. Datasäädöksen toimeenpano on vielä kesken, eikä sen tuomista mahdollisuuksista ole vielä tarkempaa selvyyttä. TYM-laki tai toisiolakikaan eivät semmoisenaan tarjoa yksiselitteistä oikeudellista edellytystä muiden toimijoiden tiedon käyttöön nimenomaisesti ennustemallien kehittämiseen. Tarvitaan poikkihallinnollinen toimija, jolla olisi mahdollisuus yhdistää eri osapuolten intressejä. Selvää vastausta tutkimuskysymykseen, kuinka tiedonjaon rajoitukset ja yhteisrekisterinpito vaikuttavat ennustemallien hyödyntämiseen TYM-

palveluiden näkökulmasta ja mitä dataa voidaan käyttää, ei opinnäytetyössä pystytty antamaan. Tiedonjakoa rajoittavia lainsäädännöllisiä seikkoja tarkasteltiin monessa osiossa, mutta tällä hetkellä yksikään lainsäädännön linjaus tai tulkinta ei anna selvää vastausta siihen mitä toimijoiden välistä dataa ennustemallien kehittämisessä on mahdollista käyttää.

Tämä työ opetti minulle, että julkisesti rahoitettujen työllistymispalvelumallien suunnittelu, toteutus ja hallinnointi on monimutkaista ja vaatii usein oppimista käytännön kautta. Kokemukset osoittavat, että monet hyödyt konkretisoituvat vasta ajan myötä. Automaattinen tiedonvaihto palveluntarjoajien ja viranomaisten välillä tehostaa järjestelmää, vähentää hallinnollista taakkaa ja parantaa läpinäkyvyyttä. Se voi myös vahvistaa tietosuojaa rajaamalla tiedonvaihdon koskemaan vain lakisääteisesti määriteltyjä tietotarpeita. Teknologioiden hyötyjen ymmärrys on usein riittämätöntä, mikä hidastaa niiden käyttöönottoa. Matalan kynnyksen testiympäristöt voisivat osoittaa, että ennustemallit tuottavat nopeasti vähintään yhtä tarkkaa tietoa, kuin ihmisten tekemät päätökset. Vaikka ennustemallit on tarkoitettu parantamaan jo olemassa olevia toimintatapoja, voi käyttöönotossa painopiste siirtyä liikaa itse teknologian käyttöön, eikä ennustemallin läpikohtaiseen ymmärrykseen. Tähän on kiinnitettävä erityistä huomiota, ettei pääse syntymään tilannetta, jossa ennustemallin tuottamaa tulosta ei osata selittää.

Työssä tutkittiin, millaisia oikeudellisia, teknisiä ja organisatorisia edellytyksiä datan jakamiseen ja yhteentoimivuuteen liittyy julkishallinnon ennustemallien käytössä. Ennustemallien kehittäminen vaatii usein henkilötason tiedon yhdistämistä esimerkiksi työllisyys-, sosiaali- ja terveyssektoreilta, mikä nostaa esiin kysymyksiä rekisterinpitäjien vastuista, oikeusperusteista, käyttötarkoituksen rajoista sekä tietosuoja-asetuksen ja tekoälyasetuksen yhteensovittamisesta. Ilman selkeitä kansallisia linjauksia voi olla haastavaa arvioida, missä määrin ja millä ehdoilla ennustemalleihin liittyvä tietojenkäsittely on sallittua tai eettisesti hyväksyttävää. Tämä korostaa tarvetta myös sääntelyllisesti tuetuille kokeiluympäristöille, kuten sääntelyhiekkalaatikoille, joissa monitoimijaiset ennustemallit voitaisiin kehittää ja testata valvotusti, yhteistyössä tietosuojaviranomaisten ja muiden ohjaavien sekä valvovien tahojen kanssa. Monen toimijan välisiä viranomaisen hallinnoimia kokeiluympäristöjä ei ole vielä kehitetty, joten ennen niiden rakentamista ei kokeiluille ole selvää kevyempää keinoa.

Vaikka tekoälyasetus (AI Act) painottaa sääntelyhiekkalaatikoiden merkitystä innovoinnin edistämiseksi, niiden käytännön toteuttaminen edellyttää organisatorisia ja hallinnollisia järjestelyjä, erityisesti julkishallinnon toimijoiden välillä. Monialaisen tuen yhteistoimintamallin näkökulmasta onnistuminen edellyttäisi, että nimetään jokin koordinaatiotaho, joka kokoaa yhteen TYM-yhteistoimintamallin palveluntuottajat, säätelviraanomaiset sekä itse teknologian kehittäjät. Koordinaatiotahon tehtävänä olisi varmistaa juridinen ennakkoharkinta, sekä tukea teknistä ja organisatorista oppimista viranomaisten kesken. Kun valvova viranomainen on mukana jo kehitysvaiheessa, voidaan vaatimustenmukaisuus todeta ajoissa ja olla ennalta perillä haasteista ja muutostarpeista. Tämä parantaa ymmärrystä tekoälyn mahdollisuuksista ja riskeistä. Sääntelyhiekkalaatikat voisivat tuoda liikkumavaraa toimintamallien simulaatioille, jolloin laillisuusvaikutuksia olisi mahdollista testata turvallisesti. Kokeilujen tulokset voisivat jopa johtaa lakimuutoksiin, jos havaitaan, että turvallisen ja hyödyllisen palvelun esteenä on vanhentunutta sääntelyä. Julkinen sektori ei voi suurten digiyhtiöiden tavoin venyttää kansainvälisten lakien rajoja, joten jos koemme voimassa olevan sääntelyn rajoittavan kehitystä, on pyrittävä vaikuttamaan sääntelyyn ennakoivasti. Kokeiluympäristöjen lisäksi on pohdittava lopullista tiedonvaihdon infrastruktuuria. Viron mallissa tiedonvaihto tapahtuu suoraan järjestelmien välillä, eikä tietoa tallenneta moneen paikkaan tarpeettomasti. Tiedon käyttö perustuu selkeään lainsäädäntöön ja viranomaisten todellisiin tarpeisiin, mikä tekee mallista hyvän esimerkin monitoimijaisesta yhteistyöstä.

Ennustemallit tarjoavat merkittävää potentiaalia TYM-mallin tueksi erityisesti varhaisen tuen tarpeiden tunnistamisessa, mutta niitä ei ainakaan nyky-lainsäädännön puitteissa voida käyttää päätöksenteon automatisointiin. Ennustemallien hyödyntäminen edellyttää ihmiskeskeistä ja moniammatillista arviointia sekä läpinäkyvyyttä ja selitettävyyttä. Suomen hajautettu tiedonvaihtoinfrastruktuuri muodostaa tällä hetkellä keskeisen esteen mallien laajalle monitoimijaiselle käytölle. Lisäksi oikeudelliset epäselvyydet, kuten GDPR:n tulkinta ja tekoälyasetuksen toimeenpano, sekä puutteet julkishallinnon koordinoinnissa hidastavat kehitystä. Vaikka teknologinen valmius on kasvamassa, tarvitaan vahvempaa strategista ohjausta, kansallista koordinaatiota ja selkeitä oikeudellisia linjauksia, jotta ennustemallit voivat tukea vaikuttavaa, ihmislähtöistä ja vastuullista palveluohjausta sosiaaliturvassa. Selvitystyön teemat ovat nousseet esiin myös yhteiskunnallisessa keskustelussa, ja Orpon hallitusohjelman sosiaaliturvauudistus digitalisaatiotavoitteineen vie kehitystä oikeaan suuntaan.

Opinnäytetyössä ei tehty vertailua ajallisesta säästöstä, koska TYM-mallin palvelutarpeen arvioinnista ei ole tarkkaa ajallista mittaustietoa. Vertailukohdetta ei ole, koska itse prosessin eri vaiheet ovat yksilölliset jokaisen asiakkaan kohdalla ja niistä ei ole eri toimijoiden välistä tilastoa. Kuten Alankomaissa on todettu ennustemallin arvion olevan kokonaisuutena tarkempi kuin asiantuntijan työ yksinään, tulisi TYM-yhteistoimintamallin ennustemallia voida verrata myös muihin tapoihin arvioida palveluntarvetta. Vertailua tulisi tehdä esimerkiksi ammattilaisen tekemään kokonaisarvioon asiakkaan tilanteesta. Näin voidaan huomata, jos malli painottaa joitakin asioita liikaa tai ohittaa tärkeitä yksilöllisiä tekijöitä. Vertailu auttaa havaitsemaan mahdolliset vinoumat ja lisää luottamusta siihen, että mallin nostamat tapaukset ovat asiakkaan oikeusturvan näkökulmasta perusteltuja ja luotettavia. Samalla rinnakkaiset arviot auttavat kehittämään mallia paremmaksi ja tunnistamaan ne tilanteet, joissa sen antama ennuste ei yksin riitä päätöksenteon tueksi. Toimenpiteellä varmistetaan ennustemallin eettisesti ja oikeudellisesti kestävä käyttö. Vertailua on mahdollista tehdä, kun käytössä on vähintään kokeiluvaiheen ennustemalli, joka yhdistää eri toimijoiden dataa.

Jatkossa olisi tärkeää tutkia, miten tietosuojasääntelyä voitaisiin soveltaa kevyemmin ennustemallien kehityksen varhaisvaiheessa ilman, että luovutaan oikeusperustan arvioinnista tai rekisteröidyn suojaamisesta. Lisäksi olisi hyödyllistä selvittää, millaisilla organisatorisilla ja teknisillä ratkaisuilla voitaisiin rakentaa yhteinen digitaalinen alusta TYM-mallin osapuolille. Olisi arvioitava onko alusta jonkinlainen yhteisrekisteri, jossa sinne viety tieto on kaikkien toimijoiden käytettävissä yhteistoimintamallin osalta vai olisiko mahdollista vain jonkinlainen lohkohtainen tiedonhaku ennustemallin käyttötarkoitusta varten. Jatkokehittämisessä tulisi tarkastella myös, miten ennustemalli voidaan rakentaa joustavaksi ja helposti päivitettäväksi muuttuvien olosuhteiden ja tarpeiden varalta. Täysin oma kysymyksensä on, riittääkö tarjolla olevien palveluiden resurssit vastaamaan ennustemallin nostamia palvelutarpeita.

Ennen yhteisten toimintamallien ja digitaalisten alustojen valmistumista tulisi kuitenkin jo ennalta varautua nopeisiin muutoksiin. Kun ennustemallien yhteinen kokeiluvaihe mahdollistuu, on tärkeää, että toimintaprosessit on suunniteltu niin, että niitä voidaan mukauttaa nopealla aikataululla uusien linjausten tai teknisten ratkaisujen voimaantullessa. Ennustemallien pilotoinneissa voidaan jo ennalta ottaa huomioon seikkoja ja kehittää ratkaisuja, jotka on mahdollista integroida kansalliseen digitaaliseen alustaan. Yhteentoimivuuden varmistaminen on oleellista ja vaikka toisten toimijoiden data

sellaisenaan ei olisi käytettävissä vielä, voidaan kuitenkin valmistautumisvaiheessa kehittää erilaisia teknisiä käytäntöjä ja selvittää datan käyttömahdollisuuksia. Hyvällä vuorovaikutuksella saadaan arvokasta tietoa muiden datasta ja toimintamalleista. Henkilöstön kouluttaminen ja osaamisen vahvistaminen tekoälyn ymmärtämisen ja uusien teknologioiden osalta valmistaa organisaatiota hyödyntämään kehitettäviä ennustemalleja myös käytännössä. Kun panostetaan teknologian hyötyjen ymmärtämiseen, voidaan tehokkaammin saavuttaa asiantuntijoiden luottamus ennustemallien käyttöön.

9 Yhteenveto

Tämä opinnäytetyö osoitti, että julkisten monitoimijaisten ennustemallien suunnittelu ja luominen on monimutkainen prosessi. Työssä etsittiin vastauksia siihen, millaisia oikeudellisia, teknisiä ja organisatorisia edellytyksiä datan jakamiseen ja yhteentoimivuuteen liittyy ja kuinka tiedonjaon rajoitukset, ja yhteisrekisterinpito vaikuttavat TYM-palveluiden ennustemallien hyödyntämiseen. Kirjallisuuskatsauksen avulla tutkittiin, mitä oppeja voidaan saada muiden EU-maiden kokemuksista, sekä miten varmistetaan ennustemallien eettinen ja oikeudellisesti kestävä käyttö monitoimijaisessa yhteistyössä.

TYM-yhteistoimintamalli on uusi toimintamalli, ja vuoden 2025 työvoimapalveluiden järjestämisvastuun uudistus voi tuoda lisäpaineita, koska kehittämistyö voi jäädä muiden muutosten varjoon. Tarkastelu osoitti, että hallittu tiedonvaihto viranomaisten välillä voisi tehostaa palvelujärjestelmää, mutta sen laajamittainen käyttöönotto edellyttää ratkaisuja erityisesti tietosuojaa ja lainsäädäntöä koskeviin kysymyksiin. Ennustemallit tarjoavat potentiaalia erityisesti varhaisen tuen tarpeiden tunnistamisessa, mutta nykyinen lainsäädäntö hidastaa mallien kehittämistä. Uusia lakeja tiedonvaihdon mahdollistamiseksi on säädetty, mutta selvää lakiperustetta ei ole toimijoiden väliselle tiedonvaihdolle ennustemallin rakentamiseen. Tuloksista ilmeni, että eettinen käyttö edellyttää mallien toiminnan olevan sekä läpinäkyvää että helposti selitettävää.

Työssä korostui tarve kansallisille linjauksille, poikkihallinnolliselle koordinaatiolle, uudenlaiselle tiedonvaihdon infrastruktuurille ja matalan kynnyksen kokeiluympäristöille. Opin tämän työn myötä, että ennustemallien käyttöä tukeva monitoimijainen tiedonjako ei ole tällä hetkellä toteutettavissa ilman uusia kansallisia linjauksia ja eri toimijoiden yhteisiä ratkaisuja. Nykyiset hajanaiset käytännöt ja epäselvä sääntely muodostavat esteen, jonka

ylittäminen vaatii sekä poliittista tahtoa että viranomaisten ja palveluntuottajien välistä tiivistä yhteistyötä. Muiden maiden mallit eivät ole sellaisenaan sovellettavissa Suomeen erilaisen infrastruktuurin vuoksi. Tulevaisuudessa TYM-yhteistoimintamallia tukevan ennustemallin luomiseen on kuitenkin hyvät edellytykset. Suomessa on tehty jo onnistuneita ennustemallikokeiluja, jotka osoittavat, että palveluntarve voidaan havaita melko varmasti. Teknologiset edellytykset ovat olemassa, tarvitaan vaan lisää selvitystyötä siitä, millä edellytyksin olisi mahdollista hyödyntää kaikkien toimijoiden dataa tarkemman ja reaaliaikaisemman ennusteen saavuttamiseksi. Yhteiskunnallinen keskustelu ja hallitusohjelma digitalisaatiotavoitteineen tukevat kehitystä.

Lähteet

Agency for Digital Government. (16.1.2018). *Agreement on digital-ready legislation*.

https://en.digst.dk/media/zpzdqvq3x/en_political-agreement-regarding-digital-ready-legislation.pdf

Agency for Digital Government. (2025). *The Digital Europe Programme*. <https://en.digst.dk/digital-transformation/the-digital-europe-programme/>

AlgorithmWatch. (2020). How Dutch activists got an invasive fraud detection algorithm banned.

AlgorithmWatch. <https://algorithmwatch.org/en/syri-netherlands-algorithm/>

Benke, K., & Benke, G. (2018). Artificial Intelligence and Big Data in Public Health. *International Journal of Environmental Research and Public Health*, 15(12), Article 12.

<https://doi.org/10.3390/ijerph15122796>

Bharosa, N., Lips, S., & Draheim, D. (2020). Making e-Government Work: Learning from the Netherlands and Estonia. Teoksessa S. Hofmann, C. Csáki, N. Edelmann, T. Lampoltshammer, U. Melin, P. Parycek, G. Schwabe, & E. Tambouris (Toim.), *Electronic Participation* (ss. 41–53). Springer International Publishing. https://doi.org/10.1007/978-3-030-58141-1_4

Bray, R., & Gray, M. (2022). *Quality Indicators for Employment Services*. ResearchGate.

https://www.researchgate.net/publication/375611432_Quality_Indicators_for_Employment_Services_Quality_Indicators_for_Employment_Services_Quality_Indicators_for_Employment_Services_ii

Brownlee, J. (2021). *Machine Learning Mastery With Python: Understand Your Data, Create Accurate Models, and Work Projects End-to-End*. Machine Learning Mastery.

Capgemini. (13.10.2021). Data ecosystems on the rise. *Capgemini*.

<https://www.capgemini.com/insights/expert-perspectives/collaborative-data-ecosystems/>

Datatilsynet. (5.7.2022). *Kommuners hjemmel til AI-profileringsværktøjet Asta*.

<https://www.datatilsynet.dk/presse-og-nyheder/nyhedsarkiv/2022/jul/udtalelse-fra-datatilsynet-kommuners-hjemmel-til-ai-profileringsvaerktoejet-asta>

Datatilsynet. (2025). *What we do*. <https://www.datatilsynet.dk/english/about-us/what-we-do>

- DigiFinland. (2024a). SOTE-tekoälyn ekosysteemi. *DigiFinland*. <https://digifinland.fi/sote-tekoalyn-ekosysteemi/>
- DigiFinland. (2024b). *Tekoäly hyvinvointialueilla: Sosiaali- ja terveydenhuollon käytötapaukset ja kansallinen edistäminen*. https://digifinland.fi/wp-content/uploads/2024/03/DigiFinland_tekoaly_loppuraportti_210324.pdf
- Digitaliserings-styrelsen. (1.7.2025). *Digital samtykke*. <https://digst.dk/nyheder/nyhedsarkiv/2025/juli/nu-bliver-det-lettere-for-borgerne-at-haandtere-deres-samtykker/>
- Eläketurvakeskus. (2018). Eläketurvakeskuksen koneoppimiskokeilu – näin se tehtiin! *Eläketurvakeskus*. <https://www.etk.fi/blogit/elaketurvakeskuksen-koneoppimiskokeilu-nain-se-tehtiin/>
- European Commission. (30.6.2025). *AI Act | Shaping Europe's digital future*. <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>
- European Parliament. (2022). *Artificial intelligence act and regulatory sandboxes*.
- Flügge, A. A., Møller, N. H., Hildebrandt, T. T., & Olsen, H. P. (2022). *Er du grøn? Algoritmer til beslutningsstøtte*.
- Hallamaa, J., & Kalliokoski, T. (2023). *Vastuun kantaminen sosioteknisen järjestelmän epäonnistumisessa*. <https://journal.fi/tatt/article/view/136352/84638>
- Hallintolaki | 434/2003 | Lainsäädäntö | Finlex*. (ei pvm.). Noudettu 6. heinäkuuta 2025, osoitteesta https://www.finlex.fi/fi/lainsaadanto/2003/434#part_2__chp_8bv20230487__sec_53ev20230487__heading
- Havinga, H., Guiaux, M., Winjhoven, M., & Dusseldorp, E. (2023). (PDF) The Work Profiler: Revision and maintenance of a profiling tool for the recently unemployed in the Netherlands. *ResearchGate*. <https://doi.org/10.1111/issr.12327>
- Hendolin, M., & Hämäläinen, H. (29.3.2022). Terveysdatan sujuva ja turvallinen käyttö. *Sitra*. <https://www.sitra.fi/julkaisut/terveysdatan-sujuva-ja-turvallinen-kaytto/>

- Ibestuur. (23.9.2024). Gemeenten lopen voor op Rijk met aantal registraties in Algoritmeregister. *iBestuur*. <https://ibestuur.nl/artikel/gemeenten-lopen-voor-met-aantal-registraties-in-algoritmeregister/>
- IBM. (2021). *What Is Machine Learning (ML)?* | IBM. <https://www.ibm.com/think/topics/machine-learning>
- Innokylä. (23.6.2025). *SOTE-tekoälyn ekosysteemi* | Innokylä. <https://innokyla.fi/fi/kokonaisuus/sote-tekoalyn-ekosysteemi>
- Kallio, A., & Kolari, J. (2023). *Tekoäly 123*. <https://www.ellibslibrary.com/reader/9789523823754>
- Kalliokoski, T. (21.2.2023). Tekoälyn vastuuta voi verrata ryhmävastuuseen. *Etairos*. <https://etairos.fi/2023/02/21/tekoalyn-vastuuta-voi-verrata-ryhmavastuuseen/>
- Kallio-Könnö, S. (2025). *SOTE-tekoälyn ekosysteemin tekoälyvisio 2035*.
- Kananen, Puolitaival, H., Harri. (2019). *Tekoäly – Bisneksen uudet työkalut*. <https://bisneskirjasto-almainsights-fi.ezproxy.hamk.fi/teos/BAXBBXATCBIED#piste:tV>
- KEHA-keskus. (2024). *Tukimateriaali työllistymistä edistävän monialaisen tuen yhteistoimintamallin verkostoille. 2024*. <https://www.keha-keskus.fi/documents/d/guest/tukimateriaali-tym-verkostoille-2024>
- Kela. (2025). *TYM-yhteistoimintamalli*. Kela. <https://www.kela.fi/yhteistyokumppanit-tyollisyyspalvelut-tym-yhteistoimintamalli>
- Kelakanava (Ohjaaja). (12.12.2024). *Datayhteistyön ja kokeilun oppeja—Kela ja Helsingin kaupungin työllisyyspalvelut 21.11.2024* [Video recording]. <https://www.youtube.com/watch?v=cVDXQbQarFQ>
- Koivisto, J., Keski-Kuha, T., Lähteenmäki, J., Sourkatti, H., Pajula, J., Antikainen, E., Kettunen, P., Vesala, J., Rankka, V., & Riekkinen, J. (2023). Monialaisen palvelukäytön ennakointi tekoälyn avulla. *THL*.
- Koivisto, J., & Tiirinki, H. (2020). *Monialaisten palvelutarpeiden tunnistamisen ja ennakkoinnin toimintamallit ja työkalut – väliraportti*.

https://www.julkari.fi/bitstream/handle/10024/139138/URN_ISBN_978-952-343-459-2.pdf?sequence=1&isAllowed=y

Kuntaliitto. (2024). *5.3 Monialaisesti toteutettavat työllistymistä edistävät palvelut ja viranomaisyhteistyö* | Kuntaliitto.fi. <https://www.kuntaliitto.fi/julkaisut/TE-palvelut2024-uudistuksen-kasikirja/5-3-monialaisesti-toteutettavat-palvelut>

Kuntaliitto. (27.5.2025). *EU:n digi- ja datasäädöksistä kunnille ja kaupungeille* | Kuntaliitto.fi. <https://www.kuntaliitto.fi/kuntajohtaminen-ja-digitalisaatio/kuntien-digikehittaminen/kuntien-digikehittamisen-lainsaadanto/EUn-digi-ja-datalainsaadanto>

Lago, C. (17.3.2021). *How Estonia is using AI to tackle unemployment – Tech Monitor*. <https://www.techmonitor.ai/digital-economy/ai-and-automation/how-estonia-using-ai-tackle-unemployment>

Laki julkisen hallinnon tiedonhallinnasta | 906/2019 | *Lainsäädäntö* | *Finlex*. (ei pvm.). Noudettu 8. heinäkuuta 2025, osoitteesta <https://www.finlex.fi/fi/lainsaadanto/2019/906>

Laki sosiaali- ja terveystietojen toissijaisesta käytöstä | 552/2019 | *Suomen säädöskokoelma* | *Finlex*. (2019). <https://www.finlex.fi/fi/lainsaadanto/saadskokoelma/2019/552>

Laki työllistymisen monialaisesta edistämisestä | 381/2023 | *Lainsäädäntö* | *Finlex*. (2025). https://www.finlex.fi/fi/lainsaadanto/2023/381#sec_16__heading

Liski-Wallentowitz, H. (2024). *”TyM/TyME”-lain ja TYP-lain tarkastelu*.

Mattila, J. (13.12.2024). Yhdessä kokeillen eteenpäin: Työdataekosysteemi. *KelaLab*. <https://medium.com/kelalab/yhdess%C3%A4-kokeillen-eteenp%C3%A4in-ty%C3%B6dataekosysteemi-d90fed92c686>

Mous, A. (5.3.2024). Kabinet wijst ACM en AP aan als toezichthouders Data Act. *VPNGids.nl*. <https://www.vpngids.nl/nieuws/kabinet-wijst-acm-en-ap-aan-als-toezichthouders-data-act/>

Mucci, T. (12.8.2024). *What Is Predictive AI?* | *IBM*. <https://www.ibm.com/think/topics/predictive-ai>

- Netherlands, S. (2025). *Microdata: Conducting your own research* [Webpagina]. Statistics Netherlands.
<https://www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research>
- Nortal. (5.7.2021). *OTT – An AI-powered success story in the public sector*. Nortal.
<https://archive.nortal.com/insights/ott-an-ai-powered-success-story-in-the-public-sector/>
- Nortal. (10.6.2022a). *X-Road: Estonia's digital backbone*. <https://nortal.com/insights/x-road-estonias-digital-backbone>
- Nortal. (29.10.2022b). *Estonian Unemployment Insurance Fund prevents unemployment with Artificial Intelligence*. <https://nortal.com/insights/estonian-unemployment-insurance-fund-prevents-unemployment-with-artificial-intelligence>
- OECD. (2025). *AI and the future of social protection in OECD countries* (42. p., OECD Artificial Intelligence Papers) [OECD Artificial Intelligence Papers]. <https://doi.org/10.1787/7b245f7e-en>
- Poikola, A., Wong, D., Lähteenoja, V., & Turpeinen, M. (10.4.2024). *Suomalaisten data-avaruuksien tilannekuva*. Sitra. <https://www.sitra.fi/julkaisut/suomalaisten-data-avaruuksien-tilannekuva/>
- Popa, D. M. (2025). Frontrunner model for responsible AI governance in the public sector: The Dutch perspective. *AI and Ethics*, 5(3), 2789–2799. <https://doi.org/10.1007/s43681-024-00596-2>
- Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: Promise and potential. *Health Information Science and Systems*, 2(1), 3. <https://doi.org/10.1186/2047-2501-2-3>
- Regulation | 2016/679 | Gdpr | EUR-Lex (2016). <https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng>
- Regulation | 2023/2854 | Data Act | EUR-Lex (2023). <https://eur-lex.europa.eu/eli/reg/2023/2854/oj/eng>
- Regulation | 2024/1689 | AI Act | EUR-Lex. Noudettu 20. elokuuta 2025, osoitteesta <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>
- RvIG. (2025). *Algoritmeregister*. RvIG. <https://www.rvig.nl/algoritmeregister>
- Salminen, A. (2011). *Mikä kirjallisuuskatsaus?*

- Salovaara, S., Leinonen, J., & Silén, M. (2021). *Tietojärjestelmien avulla kerätyn tiedon hyödyntämisen esteet sosiaalialan organisaatioiden tiedolla johtamisessa*. Finnish Journal of eHealth and eWelfare.
- Sanders, N. (2015). *Forecasting Fundamentals*. Business Expert Press.
<http://ebookcentral.proquest.com/lib/hamk-ebooks/detail.action?docID=4742536>
- Sitra. (2022). *Selvityksiä 206—Tekoälyn käyttömahdollisuudet julkisella sektorilla*.
- Sivak, L., & Stulp, G. (16.7.2024). *Intro to the CBS data*.
<https://stulp.gmw.rug.nl/prefer/posts/posts/2024-07-16-intro-csb-datasets.html>
- Smuha, N. A. (Toim.). (2025). *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence* (1. p.). Cambridge University Press. <https://doi.org/10.1017/9781009367783>
- Sosiaali- ja terveysministeriö. (2025). *Sosiaaliturvauudistus*. Sosiaali- ja terveysministeriö.
<https://stm.fi/sosiaaliturvauudistus>
- Sosiaali- ja terveysministeriö Helsinki. (2023). *Digitalisaatio työllistymisen ja osallistumisen tukena. Työ- ja toimintakykyä edistävien palvelujen digitalisoimisen tiekartta 2023–2027*.
<http://urn.fi/URN:ISBN:978-952-00-5593-6>
- Spisak, S. (2024). *Digitalisaatio, työn murros ja sosiaaliturva tietokooste*.
<https://helda.helsinki.fi/server/api/core/bitstreams/f59e0c3c-c62f-414a-a46b-bc2e8aa2d6ff/content>
- Styrelsen for Arbejdsmarked og Rekruttering. (2025). *Temaside om Beskæftigelsesreformen*.
<https://star.dk/implementering/beskaeftigelsesreform/generelle-spoergsmaal-og-svar-om-reformen>
- Suomalaisten data-avaruuksien tilannekuva. (10.4.2024). *Sitra*.
<https://www.sitra.fi/julkaisut/suomalaisten-data-avaruuksien-tilannekuva/>
- The Danish Government. (2022). National Strategy for Digitalisation—Together in the digital development. 2022. <https://en.digst.dk/media/mndfou2j/national-strategy-for-digitalisation-together-in-the-digital-development.pdf>

- Tietosuojalaki | 1050/2018 | Lainsäädäntö | Finlex.* (2018). <https://finlex.fi/fi/lainsaadanto/2018/1050>
- Trail. (30.7.2024). *EU AI Act: Risk-Classifications of the AI Regulation.* <https://www.trail-ml.com/blog/eu-ai-act-how-risk-is-classified>
- Valtioneuvosto. (2022). Valtioneuvoston selonteko: Suomen digitaalinen kompassi. *Valtioneuvoston julkaisuja.*
- Valtioneuvosto. (17.4.2024). *Kansalliset suosituksen yhdyspintojen rakentamisen alueilla.*
Valtioneuvosto. [https://tyomarkkinatori.fi/dam/jcr:1f03c26d-717b-4c50-8220-5b0865e493b9/Kansalliset%20suositukset%20yhdyspintojen%20rakentamiseen%20alueille%2017.4.2024_saavutettava%20\(1\).pdf](https://tyomarkkinatori.fi/dam/jcr:1f03c26d-717b-4c50-8220-5b0865e493b9/Kansalliset%20suositukset%20yhdyspintojen%20rakentamiseen%20alueille%2017.4.2024_saavutettava%20(1).pdf)
- Vastuu Group. (14.11.2023). *Seinäjoki ja Vastuu Group alkavat rakentaa Suomeen työllisyyspalveluiden dataverkoston.* <https://www.vastuugroup.fi/asiakastiedotteet/fi-fi/tiedotteet/seinajoki-ja-vastuu-group-alkavat-rakentaa-suomeen-tyollisyyspalveluiden-dataverkoston>
- Vastuu Group (Ohjaaja). (28.1.2025). *Data ja suosittelualgoritmi työllistymisen ja osaamisen kehittämisen tukena -webinaari* [Video recording].
https://www.youtube.com/watch?v=WB_614twcEU
- Vihalemm, T., Männiste, M., Trumm, A., & Solvak, M. (2025). Specialists and Algorithms: Implementation of AI in the Delivery of Unemployment Services in Estonia. Teoksessa P. Ahrweiler (Toim.), *Participatory Artificial Intelligence in Public Social Services: From Bias to Fairness in Assessing Beneficiaries* (ss. 97–117). Springer Nature Switzerland.
https://doi.org/10.1007/978-3-031-71678-2_5
- Vodopivec, M. (2023). *Raising the bar: Designing and implementing innovative contracted-out employment services in OECD countries* (OECD Social, Employment and Migration Working Papers No. 301; OECD Social, Employment and Migration Working Papers, Vsk. 301).
<https://doi.org/10.1787/c7a819e8-en>

Wijnhoven, M. A., Dusseldorp, E., Guiaux, M., & Havinga, H. (2023). The Work Profiler: Revision and maintenance of a profiling tool for the recently unemployed in the Netherlands. *International Social Security Review*, 76(2), 109–134. <https://doi.org/10.1111/issr.12327>

Work Profiler (Werkverkenner) | CEDEFOP. (2017, syyskuuta 15).

<https://www.cedefop.europa.eu/en/tools/resources-guidance/handbook-transferability/case-studies/work-profiler-werkverkenner>

Xiao, P. (2022). *Artificial Intelligence Programming with Python: From Zero to Hero*. John Wiley & Sons, Incorporated. <http://ebookcentral.proquest.com/lib/hamk-ebooks/detail.action?docID=6892661>

Zaken, M. van A. (16.6.2018). *Kabinet: Nederland dé digitale koploper van Europa - Nieuwsbericht - Rijksoverheid.nl* [Nieuwsbericht]. Ministerie van Algemene Zaken.

<https://www.rijksoverheid.nl/actueel/nieuws/2018/06/16/kabinet-nederland-de-digitale-koploper-van-europa>

Liite 1: Aineistohallintasuunnitelma

Opinnäytetyön aineiston kuvaus

Opinnäytetyössä käytettävä aineisto koostuu kirjallisuudesta, julkaisuista ja aiemmista tutkimuksista, jotka käsittelevät ennustemallien käyttöä julkishallinnossa. Näiden aineistojen pääasiallinen muoto on teksti, aineisto sisältää julkaisuja, kuten artikkeleita ja tutkimusraportteja.

Koska aineisto koostuu kirjallisuuskatsauksesta, ei tässä työssä ole tarvinnut kerätä henkilötietoja tai arkaluonteisia tietoja.

Aineiston tallennus ja säilytys

Tutkimusaineisto tallennetaan HAMK:n tarjoamaan salasanalla suojattuun pilvipalveluun, jossa aineisto säilytetään turvallisesti. Aineistosta tehdään varmuuskopiot kirjoittajan salasanalla suojattuun tietokoneeseen C:/ juureen, jotta aineistojen säilyvyys ja saavutettavuus voidaan varmistaa koko prosessin ajan.

Aineistoa käsittelee ainoastaan opinnäytetyön tekijä, sekä opinnäytetyön ohjaaja.

Opinnäytetyössä ei käsitellä arkaluontoista dataa, eikä henkilötietoja, joten aineisto voidaan säilyttää pilvipalvelussa.