Johannes Alaperä

KONEOPPIMISEN MAHDOLLISUUDET ESINEIDEN INTERNETIN KYBERTURVALLISUUDEN KEHITTÄMISESSÄ

JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2023

TIIVISTELMÄ

Alaperä, Johannes
Koneoppimisen mahdollisuudet esineiden internetin kyberturvallisuuden kehittämisessä
Jyväskylä: Jyväskylän yliopisto, 2023, 39 s.
Tietojärjestelmätiede, kandidaattitutkielma
Ohjaaja(t): Seppänen, Ville

Esineiden internet on noussut suureksi ilmiöksi ja se jatkaa kasvamista. Tämä kasvu tuo mukanaan myös uusia haasteita esineiden internetin kyberturvallisuudelle. Esineiden internetin kyberturvallisuuden suuret puutteet vaativat uudenlaisia ja tehokkaita ratkaisuja. Tämän kandidaattitutkielman tarkoituksena on systemaattisen kirjallisuuskatsauksen avulla tarkastella koneoppimisen algoritmien ja mallien mahdollisuuksia esineiden internetin kyberturvallisuuden kehittämiseen sekä parantamiseen ja luoda kokonaiskuva koneoppimisen sovelluksen esineiden internetin kyberturvallisuuden ratkaisuissa. Tutkielmassa esitellään erilaisia koneoppimisen menetelmiä, joilla yritetään vastata mahdollisilla ratkaisuilla esineiden internetin esitetyille kyberturvallisuushaasteille ja -ongelmille. Opinnäytetyö perehtyy myös koneopin menetelmien kohtaamiin ongelmiin niitä sovellettaessa esineiden internetin järjestelmiin ja laitteisiin. Tutkielma esittää esineiden internetin käsitteen ja sen eri kerrokset, sekä esineiden internetin kohtaamat kyberhyökkäystyypit ja ainutlaatuiset haasteet kyberturvallisuuden varmistamisessa. Koneoppimisen menetelmien käytön suurimpia kyberturvallisuusetuja esineiden internetissä ovat sen kyky käsitellä esineiden internetin tuottamaa massiivista datamäärää ja luoda siitä älykkäästi keinoja havaita tunkeutumisia esineiden internetin järjestelmissä ja ennaltaehkäistä niiden mahdollisuutta aiheuttaa laajamittaista häiriötä esineiden internetin toiminnassa, ja herkkäluonteisen sekä yksityisen datan pääsemistä hyökkääjien ja tuntemattomien osapuolten käsiin.

Asiasanat: koneoppiminen, kyberturvallisuus, esineiden internet, kyberhyökkäys

ABSTRACT

Alaperä, Johannes
The possibilities of machine learning for the development of cybersecurity for the Internet of Things
Jyväskylä: University of Jyväskylä, 2023, 39 pp.
Information Systems, Bachelors' thesis
Supervisor(s): Seppänen, Ville

The Internet of Things has become a massive phenomenon that keeps growing. This growth brings about new challenges for the cybersecurity of the Internet of Things. The vastly lacking security on the Internet of Things means new and efficient solutions are required. The aim of this bachelor's thesis is to inspect and create an overview of the possibilities to develop and improve the cybersecurity of the Internet of Things with machine learning methods through a systematic literature review. This literature review presents different machine learning methods to try to answer the challenges and problems that the Internet of Things face through possible solutions. The dissertation also delves into the problems faced when applying machine learning methods to the Internet of Things. The literature review defines the concept of the Internet of Things, its' the different layers, the different types of cybersecurity attacks that the Internet of Things faces, and the unique challenges of ensuring the cybersecurity for the Internet of Things. The greatest advantages of the methods of machine learning in the Internet of Things is how it's able to process massive amounts of the data generated by the Internet of Things and use it to create intelligent ways of detecting intrusions in Internet of Things systems and prevent the possibility of them causing widespread disturbance of the functionality of the Internet of Things as well as to prevent sensitive and private data from getting in the hands of attackers and unknown parties.

Keywords: machine learning, cybersecurity, internet of things, cyberattack

KUVIOT

Kuvio 1 Systemaattisen kirjallisuuskatsauksen vaiheet (suom. Okoli & Schabram, 2010, 9) 7
Kuvio 2 Esineiden internetin tyypillinen rakenne (suom. Wu ym., 2020, 153827) 17
Kuvio 3 Esimerkkikuvitus koneoppimisen ja syväoppimisen mahdollisesta roolista IoT-turvallisuudessa (suom. Al-Garadi ym., 2020, 153827) 29

TAULUKOT

Taulukko 1 Esineiden internetin piirteet ja niistä johtuvat ainutlaatuiset kyberturvallisuushaasteet (suom. Boukerche & Coutinho, 2021, s. 395) 21

SISÄLLYS

TIIVISTELMÄ
ABSTRACT
KUVIOT
TAULUKOT

1 JOHDANTO 6

2 KONEOPPIMINEN 8
2.1 Koneoppimisen määritelmä 8
2.2 Koneoppimisen menetelmät 9
2.3 Yleisimmät koneoppimisalgoritmit 10
2.3.1 Syväoppiminen ja sen menetelmät 12

3 ESINEIDEN INTERNETIN KYBERTURVALLISUUS 14
3.1 Esineiden internet 14
3.1.1 Esineiden internetin rakenne 15
3.2 Kyberturvallisuuden määritelmä 17
3.2.1 Kyberturvallisuuden, ICT-turvallisuuden ja tietoturvallisuuden erot 17
3.3 Esineiden internetin kyberturvallisuuden haasteet ja ongelmat 19
3.4 Esineiden internetin kyberturvallisuusperiaatteet 21
3.5 Esineiden internetin kohtaamat kyberhyökkäykset 23

4 TEKOÄLYN KÄYTTÖ ESINEIDEN INTERNETIN KYBERTURVALLISUUDEN PARANTAMISESSA 26
4.1 Koneoppimisen soveltaminen esineiden internetin kyberturvallisuudessa 26
4.2 Koneoppimisen soveltamisen haasteet esineiden internetissä 29
4.3 Tulevaisuuden näkymä koneoppimiselle esineiden internetin kyberturvallisuudelle 31

5 YHTEENVETO 34

LÄHTEET 36

1 JOHDANTO

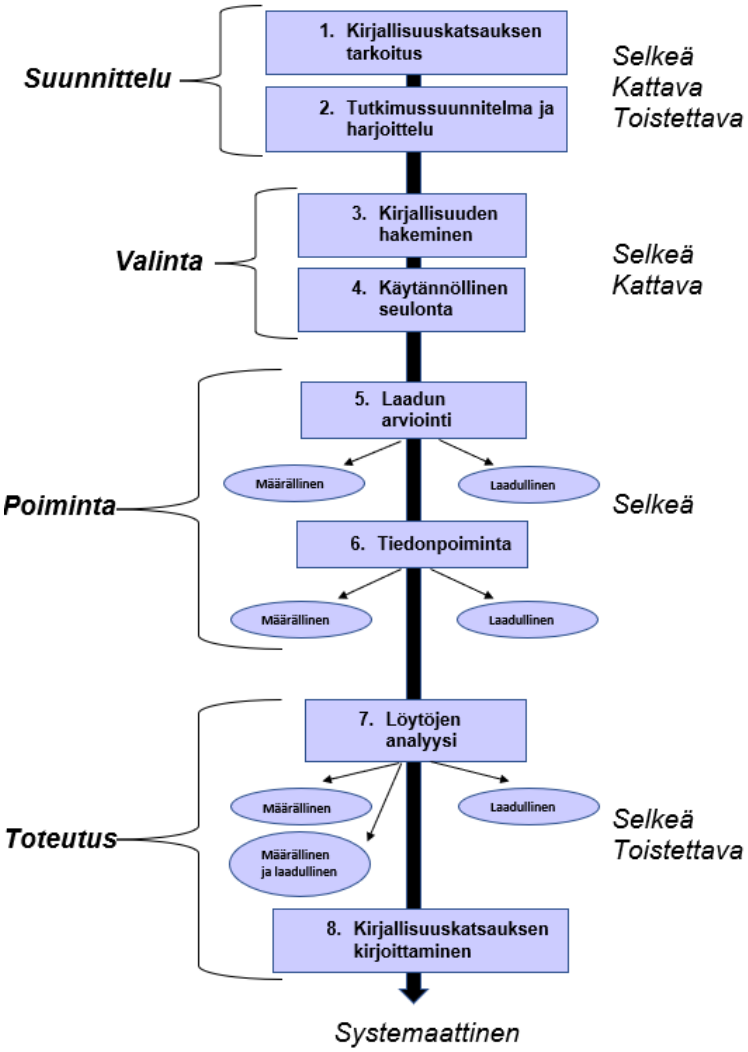
Esineiden internet (Internet of Things, IoT) on tulevaisuuden mahdollisuus yhdistää digitaalinen sekä fyysinen maailma yhteen, mutta samanaikaisesti kyberturvallisuuden uhka. Esineiden internetin eksponentiaalinen kasvu on johtanut suurempiin turvallisuus- ja yksityisyysriskeihin esineiden internetissä. Monet tämänkaltaiset riskit voidaan liittää IoT-laitteiden haavoittuvuuksiin, jotka johtuvat kyberrikollisuudesta ja järjestelmien resurssien huonosta käytöstä (Abomhara & Køien 2015). Kyberturvallisuus on jatkuva ja kehittyvä ala, jonka täytyy mukautua uusiin teknologioihin ja niiden sovelluksiin. Se, miten ja millä keinoilla esineiden internetin kyberturvallisuutta voidaan parantaa nyt ja tulevaisuudessa on tärkeä tutkimuskohde ihmisten yksityisyyden ja turvallisuuden säilyttämiseksi, sekä tietoturvallisuuden ylläpitämisen vuoksi.

Tämän opinnäytetyön tarkoituksena on tutkia koneoppimisen keinoja ja menetelmiä parantaa esineiden internetin kyberturvallisuutta. Koneoppiminen on tulevaisuuden tekoälyn käytön kannalta tärkeä tapa kouluttaa tekoälyn toimintaa. Esineiden internetin laitteista tulee nopeasti ubiikkeja, kun taas esineiden internetin (IoT) palvelut ovat muuttuneet läpitunkeviksi. Ubiikkius eli kaikkiallisuus tarkoittaa kaikkialla läsnä olevaa, joka paikkaista asiaa.

Tutkielma on tehty systemaattisena kirjallisuuskatsauksena hyödyntäen Okolin ja Schabramin (2010) 8-asteista mallia systemaattiselle kirjallisuuskatsaukselle (kuvio 1). Lähdekirjallisuuden etsiminen on toteutettu käyttämällä hakukanavoina Google Scholaria, IEEE Xploreria, JYKDOKia ja Scopusta. Tutkielman hakutermejä ovat “cyber security”, ”machine learning”, ”internet of things” sekä näiden termien yhdistelmät. Lähteet on valikoitu niiden julkaisufoorumin (lyh. JUFO) luokituksen perusteella.

Tämä tutkielma pyrkii vastaamaan seuraaviin tutkimuskysymyksiin:

- Millä keinoilla koneoppiminen voi parantaa ja kehittää esineiden internetin kyberturvallisuutta?
- Mitä mahdollisuuksia koneoppimisella on tulevaisuudessa esineiden internetin kyberturvallisuuden parantamiseen?



Kuvio 1 Systemaattisen kirjallisuuskatsauksen vaiheet (suom. Okoli & Schabram, 2010, 9)

Tutkielman rakenne on seuraavanlainen: Ensimmäinen luku, eli johdanto käsittelee tutkimuksen taustan ja tutkimusmenetelmät. Toinen luku määrittelee koneoppimisen käsitteen sekä keskittyy sen eri menetelmiin. Kolmas luku tarkastelee kyberturvallisuutta ja sen määritelmää, sekä esineiden internetin määritelmää sekä kerroksia ja esineiden internetin kyberturvallisuuden haasteita, periaatteita ja esineiden internetin kohtaamia kyberhyökkäyksiä. Neljäs luku perehtyy koneoppimisen eri menetelmien soveltamiseen esineiden internetin kyberturvallisuuden kehittämisessä sekä haasteihin koneopin soveltamisen esineiden internetissä, ja viimeisenä tarkastellaan koneopin menetelmien tulevaisuutta esineiden internetin kyberturvallisuusratkaisuina. Viimeisessä luvussa, eli yhteenvedossa, kootaan tutkimuksen tulokset ja johtopäätökset yhteen ja pohditaan jatkotutkimuskysymyksiä.

2 KONEOPPIMINEN

Tässä luvussa keskitytään koneoppimisen eri menetelmiin ja osa-alueiden määrittelyyn. Luku 2.1 esittelee koneoppimisen käsitteen, luku 2.2 koneoppimisen menetelmät, luku 2.3 yleisimmät koneoppimisalgoritmit, sekä myös syväoppimisen käsitteen ja sen menetelmät.

2.1 Koneoppimisen määritelmä

Koneoppiminen on tieteenala, joka keskittyy tekoälyn oppimiseen kehittämällä algoritmeja, jotka parhaiten kuvastavat tiettyä datajoukkoa (Choi, Coyner, Kalpathy-Cramer, Chiang & Campbell, 2020). Koneoppimisen ala on tekoälyn osa-alue. Kukkosen (2020) mukaan ”Laajasti ajateltuna tekoäly tarkoittaa sitä, että kone jäljittelee inhimillistä päättelykykyä: ongelmanratkaisua tai vaikkapa kielen ymmärtämistä. Tekoälyn voidaan ajatella olevan edistynyttä, usein koneoppimiseen perustuvaa, analytiikkaa yhdistettynä automaatioon.” Koneoppimisen ala on hyvin lähellä laskennallista tilastotiedettä, mikä myös keskittyy tietokoneiden tekemiin ennustuksiin. Alalla on myös yhteyksiä matemaattiseen optimointiin, josta se lainaa teoriaa, sovellusalueita sekä menetelmiä. Koneoppiminen voi myös olla valvomatonta ja sitä voidaan käyttää lähtötason käyttäytymisprofiilien oppimiseen ja perustamiseen erilaisille kokonaisuuksille, joilla voidaan löytää merkityksellisiä poikkeumia (Xin ym., 2018). Alpaydin (2014) esittää koneoppimisen olevan tietokoneiden ohjelmointia optimoimaan tiettyä suorituskykykriteeriä käyttäen hyväksi esimerkkidataa tai aikaisempaa kokemusta. Koneoppimisessa käytetään tilastotieteen teorioita matemaattisten mallien rakennukseen, sillä koneoppimisen perustehtävä on sama kuin tilastotieteessä, tehdä päätelmiä otoksesta (Alpaydin, 2014). Koneoppiminen käyttää siis tietokoneita simuloimaan ihmisten oppimista, jonka avulla tietokoneet voivat tunnistaa ja hankkia tietoa reaalimaailmasta ja parantaa sillä tiettyjen tehtävien tehokkuutta (Portugal, Alencar & Cowan, 2017). Samuel (1959) määrittelee koneoppimisen tieteenalaksi,

joka antaa tietokoneille kyvyn oppia itsenäisesti ilman sille suoraan tehtyä ohjelmointia. Koneoppimisen yksi aikaisimmista ja menestyksekkäimmistä sovelluksista oli 1950-luvulla Arthur Samuelin luoma tammia pelaava tietokoneohjelma (Michalski, Carbonell & Mitchell, 2013).

2.2 Koneoppimisen menetelmät

Koneoppimisalgoritmit voidaan luokitella neljään kategoriaan, joita kutsutaan koneoppimisen menetelmiksi tai lähestymistavoiksi: ohjattuun, ohjaamattomaan, puoliohjattuun ja vahvistusoppimiseen. Koneoppimisen menetelmän valinta tehdään saatavilla olevan datajoukon perusteella (Hussain, Hussain, Syed & Hossain, 2020). Ohjatussa oppimisen (supervised learning) menetelmässä määritellään tietyt tavoitetulosteet saavutettaviksi, jotka saadaan tietystä joukosta syötteitä. Tämän tyyppiselle oppimiselle annettava harjoitusdata, eli datajoukko, merkitään nimiöillä (label) luokittelua varten, jonka jälkeen tätä luokiteltua dataa, jossa on syötteet ja halutut tulosteet, käytetään algoritmin harjoittamiseen. Ohjatussa oppimisessa algoritmeille siis annetaan opetusdata eli harjoitusdata ja oikeat vastaukset eli nimiöt. Koneoppimisalgoritmin tarkoitus on opetusdatan avulla soveltaa sen oppimaa tietoa tosialliseen dataan. Ohjatun oppimisen menetelmät tallentavat harjoitusdatasta syöttöparametrien eli piirteiden (features) ja haluttujen tulosteiden, eli oikeiden vastausten, väliset suhteet. Ohjatun oppimisen menetelmä pyrkii siis automaattisesti tunnistamaan luokittelun säännöt sille annetun harjoitusdatan perusteella ja määrittämään reaalidataan useita luokkia, jonka jälkeen se pyrkii ennustamaan eri datan elementtien, kuten esineiden, yksilöiden ja kriteerien, kuulumisen tiettyyn luokkaan (Hussain ym., 2020; Boukerche & Coutinho, 2021; Portugal ym., 2017). Piirteet voivat olla luonteeltaan jatkuvia, binäärisiä tai kategorisia (Kotsiantis, 2007). Mallia, jolla piirteet luokitellaan, kutsutaan luokittimeksi.

Ohjaamattomassa oppimisessa (unsupervised learning) algoritmille tarjotaan vain syötteet ilman kohdetulosteita. Tämän tyyppisessä oppimisessa ei vaadita nimiöityä dataa, ja algoritmi voi tutkia yhtenäisyyksiä nimiöimättömän (unlabeled) datan ilmentymissä ja lajitella datan eri luokkiin (Hussain ym., 2020). Algoritmi on ohjaamaton, jos datan ilmentymien mallit ei perustu mihinkään kohdetulosteisiin vaan ne ovat algoritmin pääteltävissä (Choi ym., 2020). Ohjaamattoman oppimisen algoritmeja käytetään useimmiten etsimään datasta piilossa olevia kuvioita. Ohjaamattomassa oppimisessa algoritmeille annetaan dataa reaalimaailmasta, jonka avulla niiden tulee oppia siitä itsenäisesti. (Portugal ym., 2017). Ohjaamattoman algoritmin suurin etu on että, koska se ei tarvitse nimiöityä dataa harjoittamiseen, se vähentää kompleksisuutta ja vaadittuja resursseja. Tehokkaat ohjaamattoman oppimisen ratkaisut vaativat kuitenkin tarkkoja valintoja piirteiden suhteen (Boukerche & Coutinho, 2021).

Puoliohjattu oppiminen (semi-supervised learning) on ohjatun ja ohjaamattoman oppimisen välimuoto. Puoliohjatun oppimisen algoritmeissa vain osalle ilmentymistä on nimiö toisin kuin ohjatussa ja ohjaamattomassa, missä nimiö

joko on olemassa kaikille ilmentymille tai ei ollenkaan. Puoliohjatuissa algoritmeissa kaikilla piirteillä ei ole niitä vastaavia kohdetulosteita. Algoritmeilla on käytössä opetusdata, josta puuttuu tietoa, ja jonka avulla niiden tulee oppia. Puoliohjatut algoritmit pystyvät siis oppimaan ja tekemään johtopäätöksiä puutteellisenkin datan avulla. Käytännössä nimiöiden lisääminen datajoukkoon on kallista ja vaatii ihmisasiantuntijoiden apua, joten puoliohjatut algoritmit ovat parhaita mallinrakennukseen (Hussain ym., 2020; Portugal ym., 2017).

Vahvistusoppimisessa (reinforcement learning) ei ole ennalta määritetty tiettyjä tulosteita vaan algoritmi oppii palautteesta mitä se saa, kun se on vuorovaikutuksessa ympäristönsä kanssa. Algoritmi suorittaa jonkun toiminnon ja tekee päätöksiä sen saaman onnistumisen perusteella. Sen käyttäytyminen mallintaa ihmisten ja eläinten oppimiskäyttäytymistä (Hussain ym., 2020). Vahvistusoppimista käytetään pääasiassa päätöksentekoon. Jokaisen päätöksen lopputulemaa pidetään palkintona tai rangaistuksena, jonka perusteella arvioidaan päätöstä ja päivitetään päätöksentekotoimintoa. Oppija, tai agentti oppii kokemuksistaan yrityksien ja erehdyksien avulla (Samie, Bauer & Henkel, 2019). Agentille ei kerrota mitä sen tulee tehdä, vaan sen tulee itsenäisesti selvittää mitkä toiminnot tuottavat parhaan lopputuleman kokeilemalla jokaista eri päätöstä vuorotellen (Kotsiantis, 2007).

2.3 Yleisimmät koneoppimisalgoritmit

Yleisimpiin koneoppimisen algoritmeihin kuuluvat muun muassa k:n lähin naapuri (K-Nearest Neighbor, k-NN) tukivektorikoneet, päätöspuut, ja naiivi Bayes-luokittelija (Xin ym., 2018). Ohjattu oppiminen on yleisin menetelmä koneoppimisessa, missä tulosteet luokitellaan syötteen perusteella käyttäen oppimisdataa, mikä on oppimisalgoritmi. Ohjattu oppiminen luokitellaan luokittelu- ja regressio-oppimiseksi. Luokittelu oppiminen on ohjattu koneoppimisen algoritmi, missä tuloste on kiinteä diskreetti arvo tai luokka, kuten ”tosi”, ”epätosi” tai ”kyllä”, ”ei”. Luokitteluoppimiseen kuuluvat muun muassa tukivektorikone (Support Vector Machine, SVM), Bayesin teoreema, k-NN, satunnaismetsä (Random Forest) ja assosiaatiosääntö (Association Rule). Tukivektorikoneen algoritmia käytetään analysoimaan dataa, joka käyttää regressio- ja luokitteluanalyysiä (Tahsien, Karimipour & Spachos, 2020). Tukivektorikoneen luokittelu perustuu jakautuvan hypertason luomiseen datamääritteissä kahden tai useamman luokan välille, jotta saadaan maksimaalinen etäisyys hypertason ja jokaisen luokan vierekkäisimpien otospisteiden välille. K:n lähin naapuri on parametriton menetelmä, joka tyypillisesti käyttää euklidistä etäisyyttä etäisyysmittarina. K-NN luokitin luokittelee syötteet järjestelmälle haitallisiksi tai normaaleiksi toiminnoiksi. Järjestelmälle uudet tuntemattomat syötteet luokitellaan joko haitallisiksi tai normaaleiksi syötteen lähimpien naapureiden äänien perusteella, joita on valittu määrä. Toisin sanoen K-NN päättää tuntemattomien syötteiden luokan enemmistöäänestyksellä (Al-Garadi ym., 2020).

Bayesin teoreema perustuu tilastotieteen ehdollisen todennäköisyyden oppimisjakaumaan, jota kutsutaan Bayesilaiseksi todennäköisyydeksi. Bayesin teoreemaan pohjautuva ohjatun oppimisen menetelmä saa uusia tuloksia nykyisen tiedon perusteellä käyttäen Bayesilaista todennäköisyyttä. Tätä menetelmää kutsutaan naiiviksi Bayesiksi (Naive Bayes, NB). Naiivi Bayes on yleisesti käytetty oppimisalgoritmi, joka vaatii ennakkotietoa toimiakseen todennäköisten tulosten ennustamisessa (Tahsien ym., 2020). Naiivi Bayes luokitin on suosittu ohjatun oppimisen menetelmä sen yksinkertaisuutensa vuoksi. Naiivi Bayes laskee posteriorin todennäköisyyden ja käyttää Bayesin teoreemaa ennustamaan todennäköisyyttä, että nimiöimättömien esimerkkien tietty joukko piirteitä sopii nimenomaiselle nimiölle, sillä oletuksella, että piirteet ovat riippumattomia toisistaan (Al-Garadi ym., 2020).

Satunnaismetsä on koneoppimisen menetelmä, joka käyttää muutamaa päätöspuuta (decision tree) rakentaakseen algoritmin, jolla se luo tarkan arviointimallin tuloksille. Satunnaismetsän puita kehitetään ja koulutetaan satunnaisesti tiettyyn tarkoitukseen, josta muodostuu lopullinen tulos mallista. Vaikka satunnaismetsän luokitin rakennetaan käyttämällä päätöspuita, luokittelualgoritmit poikkeavat toisistaan huomattavasti. Satunnaismetsä ottaa huomioon tulosten keskiarvon ja vaatii vähemmän syötteitä. Päätöspuu on luonnollinen ohjatun oppimisen menetelmä, joka on nimensä mukaisesti verrattavissa puuhun, jossa on oksia ja lehtiä. Päätöspuun oksat ovat sen reunoja ja lehdet sen solmuja. Päätöspuita käytetään luokittelemaan otoksia niiden piirteiden arvojen mukaan (Tahsien ym., 2020; Al-Garadi ym., 2020). Assosiointisääntö algoritmeja käytetään tuntemattoman muuttujan tunnistukseen tutkimalla useiden erilaisten muuttujien suhdetta toisiinsa opetusdatassa. Esimerkiksi jos datajoukossa E on muuttujat A, B ja C, assosiointisäännön algoritmi tutkii näiden muuttujien suhdetta löytääkseen niiden välisiä korrelaatioita ja rakentaen siten mallin, jolla ennustetaan uusien otosten luokat (Al-Garadi ym., 2020).

Klusterointi on ohjaamattoman oppimisen menetelmä, jossa eritellään datan rakennetta ja piilossa olevia kaavoja. Klusteroinnissa luodaan rypäitä eli klusteroidaan dataa ryhmiin, joilla on samanlaisia piirteitä ja yhteinen rakenne sekä joissa datapisteet eri ryppäissä ovat erilaisia (Samie ym., 2019). Regressio-oppiminen on oppimista, jossa oppimisen tuloste on reaaliluku tai jatkuva muuttuja, riippuen sille annettavista syötemuuttujista. Regressio-oppimisen algoritmeihin kuuluvat esimerkiksi päätöspuu, neuroverkko ja kokoelmaoppiminen. Neuroverkko (neural network) on menetelmä, joka perustuu aivojen hermosolujen toimintaan. Neuroverkko algoritmeissa on yleensä kaksi neuroverkko kategoriaa: hierarkkiset ja toisiinsa kytketyt, jotka perustuvat neuronin toiminnallisiin kerroksiin. Toiminnalliset kerrokset ovat yleensä syötekerros, piilokerros ja tuloskerros. Kokoelmaoppiminen (ensemble learning) on koneoppimisen algoritmi, joka yhdistää useampien erilaisten luokittelumenetelmien tulosteet, tuottaakseen kollektiivisen tulosteen sekä parantaakseen luokittelun tehokkuutta. Koska kokoelmaoppiminen käyttää useita oppimisalgoritmeja hyväkseen, sen avulla pystytään ratkaisemaan useimpia ongelmia. Kokoelmaoppiminen on hyvin

monimutkainen verrattuna mihinkään yksittäiseen luokittelumenetelmään (Tahsien ym., 2020; Al-Garadi ym., 2020).

K:n keskiarvon klusterointi (K-means clustering) perustuu ohjaamattoman oppimisen menetelmään. Menetelmällä pyritään löytämään rypäitä datasta. K viittaa algoritmien luomien rypäiden määrään. Menetelmä toimii iteratiivisesti varaamalla jokainen datapiste yhteen k-rypääseen niiden piirteiden perusteella. Jokainen rypäs sisältää otoksia, jotka sisältävät samanlaisia piirteitä. K-keskiarvon algoritmi soveltaa iteratiivista jalostusta, jotta saadaan lopullinen tulos. Algoritmille annettavat syötteet koostuvat rypäistä (k) ja datajoukosta, joka sisältää joukon piirteitä jokaiselle otokselle datajoukossa (Al-Garadi ym., 2020).

2.3.1 Syväoppiminen ja sen menetelmät

Syväoppiminen on uusi ala koneoppimisen tutkimuksessa, joka keskittyy neuroverkkojen kehittämiseen, jotka simuloivat ihmisaivoja analyyttista oppimista varten. Se matkii ihmisaivojen mekanismeja käsitellä dataa, kuten ääniä, kuvia ja tekstejä. Koneoppimisen menetelmien tyyliin syväoppimisen menetelmissä on sekä ohjattua oppimista että ohjaamatonta oppimista. Toisin kuin koneoppimisen menetelmät, syväoppimisen menetelmissä piirteiden erottaminen on automatisoitu (Xin ym., 2018). Syväoppimisen algoritmit siis rakennetaan keinotekoisista neuroverkoista (artificial neural network, ANN tai neural network, NN) (Pantoja, Behrouzi & Fabris, 2018).

Keinotekoiset neuroverkot ovat mallinnettu biologisen neuroverkon (neuronipiirin) tavoin. Jokainen neuroverkko sisältää solmuja (vastaavat solukeskusta eli neuronin runko-osaa), jotka kommunikoivat keskenään yhteyksien välityksellä (vastaavat viejähaarakkeita ja tuojahaarakkeita neuronissa). Syväoppimista rajoittaa opetusdatan laatu ja määrä, millä sen mallia koulutetaan. Syväoppiminen kärsii lisäksi ”musta laatikko” (black box) ongelmasta, missä syöte annetaan algoritmille ja siitä saadaan tuloste, mutta ei olla varmoja mitä piirteitä voitiin tunnistaa ja miten ne vaikuttivat mallin tulosteeseen (Choi ym., 2020).

Syväoppimisen suurin etu verrattuna koneoppimiseen on sen parempi suorituskyky suurien datajoukkojen käsittelyssä. Konvoluutioneuroverkot (Convolutional neural networks) ovat yksi syväoppimisen menetelmä. Konvoluutioverkko koostuu kahdenlaisista vuorottelevista kerroksista: konvoluutiokerroksista ja yhdistämiskerroksista. Suurin etu konvoluutioneuroverkkojen käytössä on sen sovellus syväoppimisen algoritmien kouluttamiseen. Konvoluutioverkoilla voidaan opettaa algoritmeja automaattisesti tunnistamaan piirteitä raa’asta datasta korkealla suorituskyvyllä. (Al-Garadi ym., 2020). Konvoluutioneuroverkot ovat olleet erittäin tehokkaita tietyissä alueissa, kuten kuvan tunnistamisessa ja luokittelussa (Pantoja ym., 2018). Takaisinkytketyt neuroverkot (recurrent neural networks) ovat syväoppimisen algoritmien elintärkeä kategoria. Takaisinkytketyt neuroverkot käsittelevät perättäistä dataa ja niitä käytetään, koska ne pystyvät hallitsemaan peräkkäistä dataa tehokkaasti. Syvää vahvistusoppiminen (Deep Reinforcement Learning, DRL) toimii samanlaisesti kuin

vahvistusoppiminen, eli jossa algoritmi johtaa optimaalisen ratkaisun yrityksen ja erehdyksen avulla ilman aikaisempaa tietoa sen ympäristöstä (Al-Garadi ym., 2020).

3 ESINEIDEN INTERNETIN KYBERTURVALLISUUS

Tämä luku keskittyy esineiden internetin eri osa-alueisiin, siihen liittyviin kyberturvallisuusongelmiin ja toimintaperiaatteisiin. Luku 3.1 selittää esineiden internetin käsitteen ja yleisesti sen ominaisuuksia, sekä kuvailee esineiden internetin eri kerroksia, jotka muodostavat sen kokonaisuuden. Luku 3.2 määrittelee kyberturvallisuuden käsitteen ja siihen kuuluvia termejä ja erittelee ICT-, tieto- ja kyberturvallisuuden käsitteet toisistaan. Luku 3.3 perehtyy esineiden internetin yleisiin kyberturvallisuushaasteisiin ja luku 3.4 esineiden internetin kyberturvallisuuden periaatteisiin. Luku 3.5 esittelee esineiden internetin kohtaamia erilaisia kyberhyökkäyksiä.

3.1 Esineiden internet

Esineiden internetin Weber ja Studer (2016) määrittelee International Telecommunications Unionin (ITU) säädöksen mukaan. ITU:n määritelmän mukaan esineiden internet on globaali infrastruktuuri tietoyhteiskunnalle, joka mahdollistaa edistyneitä palveluita yhdistämällä fyysiset ja virtuaaliset esineet toisiinsa perustuen jo olemassa oleviin sekä kehittyviin ja yhteentoimiviin (engl. interoperable) tieto- ja viestintäteknologioihin (ITU-T Y.2060, 2012). Esineet esineiden internetissä viittaavat arkielämän objekteihin tai esineisiin, jotka vaihtelevat älykkäistä kotitalouden laitteista, kuten älykkäistä lampuista, älykkäistä sovittimista, älykkäistä mittareista, älykkäistä jääkaapeista, älykkäistä uuneista, ilmastoinnista, lämpötilasensoreista, savuhälyttimistä, IP-kameroista hienostuneempiin laitteisiin niin kuin radiotajuustunnistusta (engl. Radio Frequency Identification, RFID) käyttäviin laitteisiin, syketunnistimiin, kiihtyvyysantureihin, sensoreihin parkkipaikoilla ja kaikenlaisiin sensoreihin autoissa (Hussain ym., 2020). Esineiden internetin tarkoitus on luoda yhteenliittymä koneiden välille kytkeäkseen ne keskenään sulavasti. Siten esineiden internet ympäröi ja yhdistyy reaalimaailmaan

fyysisten laitteiden kautta, jotka sisältävät erityyppisiä sensoreita (Ahmed, Ahmed, Khan & Shah, 2020).

Hussain ym. (2020) esittävät, että esineiden internet tarjoaa lukuisia sovelluksia ja palveluita vaihdellen kriittisestä infrastruktuurista maatalouteen, armeijaan, kodinkoneisiin ja henkilökohtaiseen terveydenhuoltoon. IoT-palveluiden alat kattavat muun muassa energia-alan, kiinteistöpidon, lääketieteen alan, jälleenmyynnin alan, kuljetusalan ja tuotantoalan. IoT-verkoston suuri skaala tuo uusia haasteita, kuten IoT-laitteiden hallinnan, valtavan määrän dataa, varastoinnin, kommunikaation, laskennan sekä turvallisuuden ja yksityisyyden hallinnan haasteet (Hussain ym., 2020). Esineiden internetin laitteet keräävät ja prosessoivat dataa sekä kommunikoivat reaaliaikaisesti useiden järjestelmien kanssa, jotka muodostavat laitteisiin yhteyksiä ja valvovat laitteiden kommunikointia. Esineiden internetiä sovelletaan useisiin yhteiskunnallisesti merkittäviin alueisiin, joista esimerkkeinä älykäs terveydenhuollon teknologia, älykäs liikenne, älykäs sähköverkko ja rakennusautomaatio (Al-Garadi ym., 2020).

IoT-laitteet tallentavat kriittistä informaatiota vaihdellen äänidatasta, valon voimakkuudesta, lämpötilojen mittaamisesta, sähkönkulutuksesta, mekaanisista liikkeistä ja kemiallisista reaktioista iskujen vaikutuksiin, biologisiin muutoksiin ja maantieteelliseen sijaintiin (Al-Garadi ym., 2020). IoT-laite on ikään kuin laitteistokomponentti, mikä esineiden internetissä yhdistää kä yttäjät digitaaliseen maailmaan. IoT-laitteita kutsutaan myös älylaitteiksi ja ovat laajasti ajateltuna kodinkoneita, terveydenhuollon laitteita, ajoneuvoja, rakennuksia, tehtaita tai mikä tahansa tietoverkkoon yhdistetty ja sensoreilla varustettu laite, joka mittaa tai antaa tietoa fyysisestä ympäristöstä, toimilaite, tai sulautettu järjestelmä (Abomhara & Køien, 2015). Esineiden internetin voidaan ajatella olevan kaikenlaisten laitteiden tai esineiden kytkemistä internetiin. IoT-järjestelmä tarkoittaa teknisesti joko kiinteän tai langattoman verkon ja verkkoon yhdistettyjen laitteiden kokonaisuutta (Empirica, 2020).

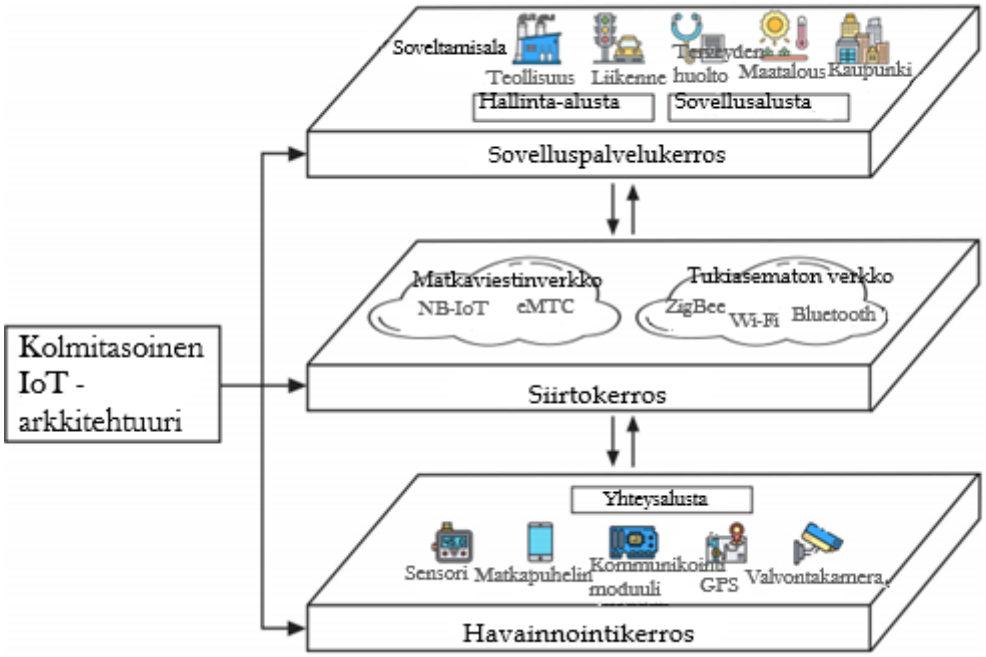
Esineiden internet on internetin evoluutiota, mutta monimutkaisempi kuin internet itsessään. Internet kommunikoi koneiden ja ihmisten välillä (eng. Machine-to-Human, M2H) kun taas esineiden internet kommunikoi koneiden kesken (eng. Machine-to-Machine, M2M) joka tuottaa älykkäitä ympäristöjä, kuten älypuhelimia, älykoteja, ja älykästä kaikkea (suom. termille smart everything) (Al Ghadeer, 2018). Esineiden internetin lopullinen tavoite on, että se pystyisi täysin hallinnoimaan itseään vastatakseen eri toimijoiden, kuten ihmisten, instituutioiden ja yritysten tarpeisiin (Lu & Xu, 2019).

3.1.1 Esineiden internetin rakenne

Esineiden internetin hierarkia tai rakenne jaetaan yleisesti kolmeen kerrokseen: sovelluskerros, verkkokerros ja havainnointikerros (Al-Garadi ym., 2020). Esineiden internet voidaan jakaa myös kyberturvallisuutta ajatellen joko neljään kerrokseen tai viiteen kerrokseen, jotka sisältävät sovelluskerroksen, verkkokerroksen ja havainnointikerroksen lisäksi väliohjelmistokerroksen ja

käyttöliittymäkerroksen. Palvelukeskeisen arkkitehtuurin perusteella puolestaan hierarkia jaetaan neljään kerrokseen, jotka sisältävät samat kolme yleistä kerrosta sekä neljännen sovelluskäyttöliittymäkerroksen (Lu & Xu, 2019). Sovelluskerros tarjoaa esineiden internetin palveluita käyttäjille mobiili- ja verkkopohjaisten ohjelmistojen kautta. Verkkokerros IoT-järjestelmässä toimii datan ja tiedon siirto- ja uudelleenohjausvälineenä käyttäen useita eri yhteyskäytäntöjä, kuten GSM:ää, LTA:ta, WiFiä, 3-5G yhteyksiä, jonka avulla se yhdistää IoT-laitteet älykkäisiin palveluihin. Havainnointikerros on ensimmäinen kerros esineiden internetin arkkitehtuurissa, joka koostuu fyysisestä kerroksesta (PHY) ja MAC-kerroksesta (medium access control) eli siirtotien varauskerroksesta. Fyysinen kerros koostuu laitteistosta, kuten sensoreista, ja siirtotienvarauskerros MAC-kerros muodostaa yhteyden fyysisten laitteiden ja verkkojen välille kommunikointia varten (Tahsien ym., 2020). Havainnointikerros on nimensä mukaisesti kerros, jossa havainnoidaan esineiden ja ympäristön fyysisiä ominaisuuksia sensoreiden, toimilaitteiden ja muiden laitteiden avulla. Havainnointiprosessi tapahtuu käyttämällä tunnistusteknologioita, kuten radiotajuustunnistusta, GPS:ää, kaksiulotteisia viivakooditunnisteita ja lukulaitteita. Tämä kerros myös muuntaa havaitun informaation digitaalisiksi signaaleiksi. Havainnoinnissa käytettävät prosessorisirut on suunniteltu ja tehty mahdollisimman pieniksi, jotta ne mahtuvat pienikokoisten IoT-laitteiden sisälle (Uprety & Rawat, 2020).

Sovelluskerros on esineiden internetin etummaisin kerros, jolla IoT-järjestelmää sovelletaan ja hallinnoidaan. Sovelluskerroksella esineiden internetin kehittäjät voivat käytännössä toteuttaa näkemyksensä IoT-järjestelmästä sille tarvittavien työkalujen avulla. Kerros hyödyntää esineiden internetin muiden kerrosten tuottamaa dataa toimiakseen. Automaattisten tunnistuslaitteiden hallinta ja IoT:n esineiden eli solmujen (node) hallinta tapahtuu tällä kerroksella. Verkkokerroksen voi ajatella olevan esineiden internetin aivot tai neuroverkko. Kerros vastaa havaintokerroksen keräämästä informaation käsittelystä. Se myös välittää informaatiota sovelluskerrokseen käyttäen langallisia ja langattomia verkkoteknologioita, kuten WiFiä, Bluetoothia, Ethernetiä, ZigBeetä, ja 3G-verkkoa. IoT-sensoreiden kerätessä valtavaa määrää dataa, se tarvitsee väliohjelmiston millä prosessoida sitä, johon käytetään pilvipalveluja pääasiallisena teknologiana (Uprety & Rawat, 2020).



Kuvio 2 Esineiden internetin tyypillinen rakenne (suom. Wu ym., 2020, 153827)

3.2 Kyberturvallisuuden määritelmä

Suomen kyberturvallisuusstrategia määrittelee kyberturvallisuuden ”tavoitetilaksi, jossa kybertoimintaympäristöön voidaan luottaa ja jossa sen toiminta turvataan” (Puolustusministeriö, 2013). Jansson ja Sihvonen (2018) määrittelevät Yhdysvaltain presidentin George W. Bushin linjauksen Comprehensive National Cybersecurity Initiative eli CNCI:n mukaan (The White House 2008) kybertoimintaympäristön olevan kokonaisuus useista toisiinsa kytköksistä olevista tietoverkoista, joissa tieto siirtyy sähköisessä muodossa koneelta ja sen käyttäjältä toiselle, tietokoneista ja tietoliikennetekniikasta, sekä erilaisia tehtäviä käsittelevistä datasäilöistä, palvelimista ja reitittimistä. Janssonin ja Sihvosen (2018) mukaan Sessions (2014) määrittelee lisäksi ihmisen osana kybertoimintaympäristöä, sillä ihminen on vielä vastuussa verkon ylläpitämisestä ja sen toiminnan olosuhteista.

3.2.1 Kyberturvallisuuden, ICT-turvallisuuden ja tietoturvallisuuden erot

Kyberturvallisuuden, tietoturvallisuuden sekä ICT-turvallisuuden käsitteitä käytetään usein synonyymeinä toisilleen. Xinin ym. (2018) mukaan kyberturvallisuus on joukko teknologioita ja prosesseja, jotka on suunniteltu suojelemaan tietokoneita, tietoverkkoja, ohjelmia sekä dataa hyökkäyksiltä ja valtuudettomalta pääsyltä, muutoksilta tai tuholta. Suomen puolustusministeriön (2013) mukaan tietoturvallisuus tarkoittaa ” järjestelyjä, joilla pyritään varmistamaan

tiedon käytettävyys, eheys ja luottamuksellisuus.” International Telecommunications Union (2008) määrittelee kyberturvallisuuden tärkeimmiksi turvallisuustavoitteiksi saatavuuden, eheyden, sisältäen mahdollisesti myös oikeellisuuden ja kiistattomuuden, sekä luottamuksellisuuden. Kyberturvallisuuden alaisuuteen kuuluvat myös valtuudeton pääsy tietoihin sekä hyökkäykset, jotka johtavat palveluiden saatavuuden keskeytykseen (Lu & Xu, 2019). ICT eli tieto- ja viestintäteknologian turvallisuus tarkoittaa teknologiapohjaisten järjestelmien suojelua, jossa tietoa tallennetaan tai välitetään. Koska tietoturvallisuus käsittää taustalla olevien informaation resurssien suojelun, voidaan sano ICT-turvallisuuden olevan yksi tietoturvallisuuden osa (Von Solms & Van Niekerk, 2013). Janssonin ja Sihvosen (2018) mukaan kyberturvallisuuden kokonaisuus käsittää enemmän asioita kuin tietoverkko- tai tietoturvallisuus kattavat.

Kyberturvallisuus tapahtuu kybertoimintaympäristössä. Kybertoimintaympäristölle tyypillisiä piirteitä ovat Puolustusministeriön (2013) mukaan ”elektroniikan ja sähkömagneettisen spektrin käyttö datan ja informaation varastointiin, muokkaamiseen ja siirtoon viestintäverkkojen avulla”. Tähän kuuluu lisäksi datan ja informaation hallinnan olennaiset fyysiset rakenteet (Puolustusministeriö 2013). Kyberturvallisuuteen liittyy kriittisen infrastruktuurin käsite. Puolustusministeriö (2013) määrittää kriittisen infrastruktuurin sellaisina toimintoina ja rakenteina, jotka ovat tarpeellisia elintärkeille toiminnoille yhteiskunnassa, mukaan lukien fyysiset laitokset ja rakenteet sekä sähköiset palvelut ja toiminnot, joita yhteiskunta käyttää ja tarvitsee.

Von Solmsin ja Van Niekerkin (2013) mukaan tieto ja tieto- sekä viestintäteknologia kyberturvallisuuden näkökulmasta ovat haavoittuvuuksien taustalla oleva syy. Kyberturvallisuuden merkittävin yksittäinen piirre on se, että kaikkea kyberturvallisuuden alaista suojeltavaa omaisuutta tulee suojella, koska haavoittuvuudet johtuvat tieto- ja viestintäteknologian, joka muodostaa kybertoimintaympäristön perustan, käytöstä. Nämä haavoittuvuudet ulottuvat myös aineettomiin omaisuuksiin. Kyberturvallisuuden tulee suojella enemmän kuin pelkästään ihmisten tai organisaation omistamaa tietoa tai tietojärjestelmien resursseja. Samalla tavoin kuin tietoturvallisuuden käsite laajentaa ICT-turvallisuuden käsitettä, jotta voidaan turvata tieto itsessään, kyberturvallisuus tulee käsittää tietoturvallisuuden laajennuksena. Kyberturvallisuus käsittelee lisäksi eettistä ulottuvuutta, sillä ongelmat, kuten nettikiusaaminen, ulottuvat lain ulkopuolelle ja muodostavat eettisen ongelman yhteiskunnalle. Eettinen ulottuvuus käsittää muun muassa myös bottiverkostot. Voidaan siis sanoa, että kyberturvallisuus on kyberympäristön suojelemista, mikä tarkoittaa digitaalista tietoa, tieto- ja viestintäteknologiaa, joka tukee kyberympäristöä, sekä kyberympäristön käyttäjien yhteiskunnallista, valtiollista ja henkilökohtaista omaisuutta, joka on uhanalainen kyberympäristössä tapahtuville hyökkäyksille (Von Solms & Van Niekerk, 2013).

3.3 Esineiden internetin kyberturvallisuuden haasteet ja ongelmat

Verkkoturvallisuus on välttämättömyys massiivisen internetin käytön seurauksena. Symantecin (2019) raportin mukaan IoT-laitteet kokevat keskiverrolta 5,200 hyökkäystä kuukautta kohti. Esineiden internetin kasvun seurauksena, tietoturvallisuusriskit lisääntyvät eksponentiaalisesti. IoT-laitteiden ominaisuuksista ja tietoliikenneprotokollista johtuen esineiden internetin tietoverkot ovat haavoittuvaisempia kuin perinteinen tietoverkko. Nämä haavoittuvaisuudet ilmenevät muun muassa virusten, tunkeutumisyritysten sekä muunlaisten kyberhyökkäysten avulla (Da Costa ym., 2019). Moneen haasteeseen, kuten tietojen salassapitoon, luottamuksellisuuteen, tietojen eheyteen, todentamiseen ja käyttöoikeuksien jakamiseen täytyy vastata ennen kuin esineiden internetin kyberturvallisuutta voidaan pitää korkeana. Useat eri teknologiat, standardointi, sekä muut parhaillaan meneillään olevat tutkimukset pyrkivät täyttämään esineiden internetin kyberturvallisuuden korkeat vaatimukset (Lu & Xu, 2019). Wu, Han, Wang ja Sun (2020) selvittivät, että esineiden internet kohtaa neljä kriittistä kyberturvauhkaa: laitteiden todentaminen, palvelunesto hyökkäykset, tunkeutumisen havaitseminen ja haittaohjelmien havaitseminen. Perinteiset kyberturvallisuusratkaisut näihin uhkiin ovat puutteellisia, koska ne eivät pysty käsittelemään suuria datajoukkoja ja niillä on huono reaaliaikainen toimintakyky sekä alhainen tehokkuus. Suurinta osaa näistä ratkaisuista ei pystytä soveltamaan esineiden internetiin. Koneoppimisen edustamat tekoälymenetelmät voivat hyödyntää valtavaa IoT-laitteiden keräämää dataa, osoittaa datasta hyödyllistä tietoa sekä tehdä datan avulla ennustuksia tuntemattomista tapahtumista, tarjoten uusia ratkaisuja näihin kyberturvallisuuden ongelmakohtiin (Wu ym., 2020).

IoT-järjestelmät perustuvat kahteen komponenttiin; järjestelmän laitteistoon ja järjestelmäohjelmistoon. Molemmissa näissä on usein suunnitteluvirheitä. Laitteiston haavoittuvuudet ovat vaikeasti havaittavissa ja myös vaikeasti korjattavissa, sillä vaikka haavoittuvuudet löydettäisiinkin, laitteiston yhteensopivuus ja yhteentoimivuus sekä korjaamisen suuri vaiva aiheuttavat ongelmia. Järjestelmäohjelmiston haavoittuvuuksia löytyy käyttöjärjestelmistä, sovellusohjelmistoista ja ohjausohjelmistoista kuten tiedonsiirtoprotokollista ja laiteohjaimista (Abomhara & Køien, 2015). Esineiden internetiin kohdistuu Al Ghadeerin (2018) mukaan kolme erilaista hyökkäystyyppiä: fyysiset hyökkäykset, ohjelmistoon kohdistuvat hyökkäykset ja IoT:n tietoverkkoon kohdistuvat hyökkäykset. Esineiden internet tuottaa valtavan määrän arvokasta dataa, ja jos tätä dataa ei välitetä ja analysoida turvallisesti, saattaa tapahtua kriittinen tietoturvaloukkaus. IoT-järjestelmän jokaisen osan syöttödataa voidaan kerätä ja tutkia, normaalien vuorovaikutusmallien määrittämiseksi, jolloin haitallinen käyttäytyminen pystytään tunnistamaan jo varhaisessa vaiheessa (Al-Garadi ym., 2020).

Abomhara ja Køien (2015) kuvaavat syitä, miksi esineiden internetiin yhdistyneet laitteet ovat erittäin arvokkaita kyberhyökkääjille: Suurin osa IoT-laitteista toimivat ilman ihmisten valvontaa, joten hyökkääjien on helppo päästä

niihin fyysisesti käsiksi. Suurin osa IoT-komponenteista kommunikoi langattomien verkkoyhteyksien kautta, jonka avulla hyökkääjä voi saada käsiinsä luottamuksellista tietoa salakuuntelemalla. Suurin osa IoT-komponenteista ei tue monimutkaisia suojausjärjestelmiä johtuen niiden matalasta virrankäytöstä ja alhaisista laskentaresursseista. Salim, Shailendra ja Parkin (2020) mukaan syitä miksi esineiden internetiin kohdistetaan palvelunestohyökkäyksiä (Denial-of-Service, DoS) ovat esimerkiksi jatkuvasti internetiin yhteydessä olevat laitteet, puute minkäänlaisista perusturvaprotokollista, sillä monien laitteiden valmistajat eivät asenna tarpeellisia turvallisuusprotokollia IoT-laitteisiin, helposti hyväksi käytettävät salasanat johtuen siitä etteivät laitteiden omistajat tyypillisesti vaihda salasanojaan, kykenemättömyys nollata valtuutusta kun hyökkääjä on ottanut haltuunsa kohdelaitteen turvallisuuteen liittyvät valtuustiedot, puute turvallisuuteen liittyvistä laiteohjelmistopäivityksistä johtuen valmistajien puutteellisesta valtuustietojen seurannasta ja tietoturvapäivitysten varmistamisesta, sekä kustannustehokkaista hyökkäyksistä. Hyökkääjien ei tarvitse ylläpitää ja sijoittaa rahaa kalliisiin serveritietokoneisiin, jolla toteuttaa hyökkäyksiään, vaan he voivat ottaa haltuunsa turvattomia IoT-laitteita halvalla tai lähes ilmaisesti verrattuna perinteisiin palvelunestohyökkäyksiin (Salim ym., 2020). Al Ghadeerin (2018) mukaan palvelunestohyökkäyksiä voidaan pitää haitallisimpina hyökkäyksinä esineiden internetiä kohtaan, sillä se voi vaikuttaa jokaiseen esineiden internetin kerrokseen.

Boukerche ja Coutinho (2021) loivat havainnollistavan luokittelun esineiden internetin erikoispiirteistä johtuvista kyberturvallisuushaasteista (taulukko 1). IoT-laitteiden heterogeenisyys on ongelma, sillä yhteysprotokollat tulee suunnitella toimimaan kaikkien laitteiden kanssa, joilla on useita erilaisia kykyjä muodostaa yhteyksiä, useita monimutkaisuuksia, myyjiä ja julkaistuja versioita. IoT-laitteet käyttävät lisäksi erilaisia teknisiä yhteenliittymiä erilaisten bittinopeuksien kanssa, jotka ovat suunniteltu tiettyihin toimintoihin. Skaalautuvuus on lisäksi ongelma, sillä joka päivä uusia laitteita kehitetään toimimaan esineiden internetin tietoverkossa. Se tarkoittaa haasteita, jotka liittyvät esimerkiksi laitteiden todentamiseen, tiedonhallintaan ja palveluidenhallintaan (Nazir, Sholla & Bashir, 2019). Lisäksi minkä tahansa suojausalgoritmin soveltaminen esineiden internetiin tuo omia haasteitaan. Butunin, Österbergin ja Songin (2018) mukaan Balte, Kashid ja Patil (2015) määrittävät näitä haasteita olevan muun muassa IoT-laitteiden heterogeenisyys ja monimuotoisuus laitteiden yhteyskäytäntöjen sekä tietojenkäsittelyalgoritmiensa puolesta, skaalautuvuus eli miljoonien eri laitteiden hallinnan ja ylläpidon haastavuus, erilaiset kommunikaatioteknologiat kuten Bluetooth ja ZigBee IoT-laitteissa, energiankulutuksen rajoitteet IoT-laitteissa, käyttäjien tietosuoja esimerkiksi siinä tapauksessa, että IoT-laite jakaa toimintatilassaan sijaintiaan verkon ylläpitäjälle tai viereisille laitteille niiden pyytäessä, itsetajunta eli IoT-laitteiden tulisi organisoitua itsenäisesti, jotta ne voivat suorittaa tiettyjä ennalta määrättyjä tehtäviä reagoidakseen todellisiin tilanteisiin ilman ihmisten puuttumista niiden toimintaan, sekä yhteentoimivuus, sillä jotta IoT-laitteet voivat kommunikoida ja jakaa dataa keskenään, tulisi laitteiden keskinäisen tiedonsiirtoformaatin olla ennalta määrätty ja standardoitu.

Taulukko 1 Esineiden internetin piirteet ja niistä johtuvat ainutlaatuiset kyberturvallisuushaasteet (suom. Boukerche & Coutinho, 2021, s. 395)

Esineiden Internetin ominaispiirteet	**IoT-sovellusten kyberturvallisuushaasteet**
Massiivinen käyttöönotto	• Data jakaantuu useamman laitteen kesken. • Laitekohtaisen turvallisuuden varmistaminen. • Tietoverkon yleiset resurssit.
Heterogeenisyys	• Laitteilla on heterogeeniset ominaisuudet. • Tarve erilaisille ratkaisuille turvaamaan erityyppisiä laitteita.
Dynaamiset verkkotopologiat	• IoT-verkkotopologia muuttuu usein johtuen hallittavissa ja hallitsemattomista olevista tekijöistä. • Topologian muutokset vaikuttavat IoT-laitteiden viestintäkaavoihin. • Tunnisteihin perustuvaa todentamista hyödyntävien kyberturvallisuusratkaisuiden tulee ottaa huomioon tietoliikenteen kaavojen muutokset.
Matalavirtainen ja matalakustanteinen viestintä	• IoT-laitteilla on huomattavia virtarajoitteita. • Tietoverkkoyhteyskäytänteet eivät toteuta vahvaa mekanismia luotettavaa viestintää varten. • Hajautetuiden kyberturvallisuusratkaisuiden tulisi harkita alhaisen luotettavuuden viestintää IoT-sovelluksissa.
Alhaisen latenssin viestintä	• IoT-sovelluksilla saattaa olla aikarajoitteita. • Monimutkaiset kyberturvallisuusratkaisut aiheuttavat lisäviiveitä.

3.4 Esineiden internetin kyberturvallisuusperiaatteet

Tietoturvan varmistaminen on yksi tärkeimmistä kyberturvallisuuden alaisuuteen kuuluvista ongelmista. Suurimpia tietoturvan ongelmia ovat tiedon luottamuksellisuus, tietosuoja ja tiedon eheys (Lu & Xu, 2019). Minkä tahansa kyberturvallisuusratkaisun tulee ottaa huomioon luottamuksellisuus, eheys ja saatavuus (Sadique, Rahmani, & Johannesson, 2018). Näiden lisäksi esineiden internetin kyberturvallisuudessa on myös tärkeää huomioida autentikaatio eli todentaminen, oikeuksien valtuuttaminen eli valtuutus sekä kiistattomuus (non-repudiation). Luottamuksellisuus on elintärkeä turvallisuusominaisuus IoT-järjestelmissä. Esineiden internetin laitteet saattavat tallentaa ja siirtää arkaluontoista tietoa, kuten lääketieteellistä tai henkilökohtaista tietoa, joka on pidettävä salassa luvattomilta yksilöiltä (Al-Garadi ym., 2020). Lisäksi sotilaallista tietoa, yksityistä kaupallishallinnollista tietoa sekä käyttäjien turvatunnuksia on pidettävä turvassa luvattomilta osapuolilta. Tiedon eheys tulee varmistaa, sillä IoT-laitteiden tuottama data siirtyy tyypillisesti langattoman tiedonsiirron avulla, ja sitä ei tulisi

pystyä muuttamaan kukaan muu kuin valtuutettu taho. Eheyden suojaus on haastavaa, sillä ainoastaan sallituilla käyttäjille tulisi olla pääsy käyttäjien erilaisiin henkilötietoihin. Lisäksi tiedot tulee hävittää turvallisesti, kun niitä ei enää tarvita. Eri IoT-järjestelmillä on useita eheysvaatimuksia, esimerkiksi potilaiden etävalvontajärjestelmillä on korkeat vaatimukset eheyden varmistamiselle satunnaisvirheiden välttämiseksi, sillä potilastieto on arkaluonteista (Abomhara & Køien, 2015; Sadique ym., 2018). Jos eheyden tarkastaminen on puutteellista, IoT-laitteiden muistiin tallennetun datan muuttaminen on mahdollista, mikä voi vaikuttaa fyysisten laitteiden tärkeimpiin toiminallisiin ominaisuuksiin pidemmän aikaa ilman, että sitä havainnoidaan. Eheysominaisuudet ovat perusteellisia, jotta voidaan taata tehokas tarkistusmekanismi, joka havaitsee minkä tahansa muutoksen epäluotettavan langattoman tietoverkon kommunikaatiossa (Al-Garadi ym., 2020).

Saatavuus on perusominaisuus onnistuneiden IoT-järjestelmien toiminnassa, sillä esineiden internetin tarjoamien palveluiden täytyy aina olla saatavilla sen valtuutetuille tahoille. IoT-laitteiden tulee aina olla saatavilla, jotta ne voivat tarkkailla ja kerätä dataa. Eritoten reaaliaikaisissa seurantajärjestelmissä esineiden internetin laitteiden saatavuus tulee olla korkea. Saatavuusvaatimukset vaihtelevat järjestelmittäin, esimerkiksi terveydenhuollon valvontajärjestelmillä tai palonvalvontajärjestelmillä on mitä todennäköisemmin korkeammat saatavuusvaatimukset kuin tienvarsien saasteantureilla. Monet kyberhyökkäykset voivat estää IoT-järjestelmien ja laitteiden saatavuuden, kuten palvelunestohyökkäykset, joten jatkuvan IoT:n palveluiden saatavuuden ylläpitäminen on kriittistä esineiden internetin turvallisuudessa. IoT-laitteiden eri laitteisto- ja ohjelmistokomponenttien tulee olla kestäviä, jotta palveluiden saatavuus pysyy myös epäsuotuisissa tilanteissa tai tilanteissa, joissa jokin haitallinen taho on läsnä (Al-Garadi ym., 2020; Sadique ym., 2018; Abomhara & Køien, 2015).

Jokaisen olion identiteetti tulisi varmistaa ennen minkään prosessin suorittamista. Koska IoT-laitteet käsittelevät ja tuottavat arkaluonteista tietoa, niiden tulee todentaa itsensä verkossa ottaakseen vastaan ja lähettääkseen eteenpäin tietoa. Kuitenkin johtuen IoT-järjestelmien luonteesta, todentamisen vaatimukset vaihtelevat eri järjestelmien kesken ja järjestelmien eri todennusvaatimukset edellyttävät erilaisia ratkaisuja. Esimerkiksi sellaisen IoT-järjestelmän tapauksessa, missä IoT-palvelu tarjoaa vahvaa turvallisuutta suuren joustavuuden sijaan, tulisi myös järjestelmän todentamisen olla vahva. Joidenkin todennusratkaisujen tulee olla vahvoja, esimerkiksi kun todennetaan pankkikortteja tai pankkijärjestelmiä. Kompromissit ovat suuri haaste kehitettäessä tehokasta todentamisjärjestelmää, kompromissi voi olla tehokkaan todennusjärjestelmän ja akkupohjaisten laitteiden välillä. IoT-järjestelmä vaatii tehokkaan todennuksen, joka pystyy tasapainottamaan järjestelmän rajoitteet ja tarjoamaan vankkoja suojamekanismeja. Valtuutukseen kuuluu IoT-järjestelmään pääsyn antaminen käyttäjille, esimerkiksi fyysiseen sensorilaitteeseen. Valtuutusominaisuus sallii vain valtuutettujen käyttäjien, todennettujen olioiden, suorittaa tiettyjä toimintoja verkossa. Käyttäjät voivat olla sekä ihmisiä, koneita että palveluita. Ainoastaan valtuutetuilla käyttäjillä tulee olla lupa päästä käsiksi sensoreiden keräämään

dataan, eli toiminto tulee suorittaa vain, jos sen pyynnön esittäjällä on tarvittava valtuutus. Valtuutus on yhtä tärkeä kuin todentaminen. Esineiden internetin laitteiden tulee pystyä lukemaan ja muuttamaan tietoa vain tietyssä osassa tietokantaa. Kyberhyökkääjät voivat saada lukuoikeuden ja oikeuden muuttaa tietokannan tietoja turvallisuudeltaan vaarantuneen IoT-laitteen kautta. Suurin haaste esineiden internetin toimintaympäristön valtuuttamisessa on, kuinka onnistuneesti myönnetään pääsy sellaisessa ympäristössä, missä ihmisten lisäksi myös fyysiset sensorit eli esineet tulee valtuuttaa olemaan esineiden internetin järjestelmän kanssa vuorovaikutuksessa (Sadique ym., 2018; Al-Garadi ym., 2020; Abomhara & Køien, 2015).

Yleisesti monissa IoT-järjestelmissä kiistattomuuden ominaisuutta ei pidetä avainominaisuutena esineiden turvallisuudessa. Sen periaate on tarjota pääsy lokitietoihin, jotka toimivat todisteena sellaisissa tilanteissa, missä laitteet tai käyttäjät eivät voi kieltäytyä toiminnosta. Kiistattomuus on tärkeä tietyissä tilanteissa, kuten maksutapahtumissa, jossa kumpikaan osapuoli, käyttäjät tai palveluntarjoajat, eivät voi perua maksutoimintoa (Al-Garadi ym., 2020; Abomhara & Køien, 2015).

3.5 Esineiden internetin kohtaamat kyberhyökkäykset

Esineiden internetin luonteen takia sen laitteet kohtaavat monenlaisia eri kyberhyökkäyksen tyyppejä. Kyberhyökkäykset esineiden internetissä viittaavat uhkaan, joka kohdistuu langattomassa verkossa IoT-laitteisiin, jossa hakkeroidaan järjestelmää käyttäjän tietoja manipuloidakseen. Kyberhyökkäykset voidaan jakaa kahteen kategoriaan: passiiviset ja aktiiviset hyökkäykset (Tahsien ym., 2020). Passiivisia hyökkäyksiä tehdään tavalla, mikä tekee niiden havaitsemisesta käytännössä mahdotonta millään keinoilla. Passiiviset hyökkäykset kohdistuvat pääosin tiedon luottamuksellisuutta kohtaan, sillä hyökkääjä yrittää pysyä piilossa ja kerätä tietoa salakuuntelemalla linjoja. Langattomat verkot ovat tyypillisesti helpompia kohteita tämän tyyppisille hyökkäyksille. Aktiiviset hyökkäykset kohdistetaan sekä tiedon eheyttä että tiedon luottamuksellisuutta kohtaan. Aktiiviset hyökkäykset tähtäävät esimerkiksi verkon kommunikoinnin häirintään, resurssien varaamiseen sekä valtuudettomaan pääsyyn. Tämän tyyppisissä hyökkäyksissä hyökkääjä vaikuttaa suoraan hyökkäyksen kohteena olevan tietoverkon toimintaan. Palvelunestohyökkäykset kuuluvat muun muassa aktiivisiin hyökkäyksiin (Butun ym., 2018). Abomhara ja Køien (2015) kuvaavat tyypillisiä kyberhyökkäyksiä mitä esineiden internet voi kohdata, mukaan lukien fyysiset hyökkäykset, tiedusteluhyökkäykset (reconnaissance attack), palvelunestohyökkäykset, käyttöoikeushyökkäykset (access attack) sekä hyökkäykset käyttäjän yksityisyyttä kohtaan. IoT-toimintaympäristön ollessa hajautettu ja valvomaton, sen laitteita, joista suurin osa toimii tyypillisesti ulkoilmaympäristössä, kuten sensorit, kohtaan voidaan tehdä myös fyysisiä hyökkäyksiä, joissa yritetään peukaloida IoT-laitteita eli IoT-järjestelmän laitteiston komponentteja.

Tiedusteluhyökkäykset ovat luvattomia järjestelmien, palveluiden tai heikkouksien kalastelua ja kartoitusta (Abomhara & Køien, 2015).

Tiedusteluhyökkäys voi olla esimerkiksi verkkoporttien skannaamista, salasanojen haistelijoiden eli snifferien käyttämistä, tietoverkkoliikenteen analyysia tai IP-osoitteen informaation tiedustelua. Palvelunestohyökkäys on tapa saada laitteen tai verkon resurssit pois sen tarkoitetuilta käyttäjiltä. Johtuen IoT-laitteiden alhaisista muistitoiminnoista ja rajoitetusta laskentaresursseista, suurin osa laitteista esineiden internetissä ovat haavoittuvaisia resurssien varaamiseen perustuville hyökkäyksille (Abomhara & Køien, 2015). Palvelunestohyökkäyksellä keskeytetään järjestelmän toiminta luomalla useita turhia palvelinpyyntöjä järjestelmän palvelimille tai verkolle, jonka takia käyttäjä ei voi käyttää IoT-laitetta tai kommunikoida sen kanssa. DoS-hyökkäykset lisäksi pitävät IoT-laitteen aina päällä, mikä kuluttaa sen akun elinaikaa. Hajotetut palvelunestohyökkäykset koostuvat useista hyökkäyksistä käyttäen eri IP-osoitteita, jolla voidaan luoda useita eri palvelinpyyntöjä palvelimille ja täten ne jumittavat palvelimen toimintaa. Esimerkiksi bottiverkko nimeltään Mirai aiheutti vahinkoa tuhansille eri IoT-laitteille häiritsemällä niiden toimintaa (Tahsien ym., 2020).

Hajautettu palvelunestohyökkäys on yritys sulkea kohteena oleva serveri joko kokonaan tai osittain keskeyttämällä normaali verkkoliikenne kohteen serverille tai tietoverkolle. Pääsijainen tarkoitus hajautetuissa palvelunestohyökkäyksissä on estää kohteen verkon tai kaistanleveyden resurssit, jotta uhrit eivät pääse palveluun käsiksi (Salim ym., 2020). Käyttöoikeushyökkäyksillä luvattomat henkilöt pyrkivät saamaan käyttöoikeuden tietoverkkoon tai laitteisiin, joihin heillä ei ole lupaa päästä käsiksi. Käyttöoikeushyökkäyksiä on kahdenlaisia, fyysinen pääsy ja etäyhteys. Fyysisessä pääsyssä hyökkääjä pyrkii pääsemään käsiksi fyysiseen laitteeseen. Etäyhteyshyökkäyksessä laitteiseen pyritään saamaan pääsy IP-yhteydellä yhdistettyihin laitteisiin. Hyökkäyksiä käyttäjien yksityisyyteen on monenlaisia ja yksityisyyden suojelusta esineiden internetissä on tullut haastavampaa johtuen suurista määristä tietoa, joka on helposti saatavilla etäyhteyden avulla. Yleisimmät hyökkäykset yksityisyyttä kohtaan ovat tiedon louhinta, kybervakoilu, salakuuntelu, laitteiden seuranta ja salasanaan perustuvat hyökkäykset. Tiedon louhinnalla tunkeutuja löytää tietoa, johon tietyt tietokannat eivät osaa ennakoida. Kybervakoilulla tunkeilija käyttää hakkerointimenetelmiä tai haittaohjelmia vakoillakseen tai saadakseen salaista tietoa henkilöistä, organisaatioista tai valtioista (Abomhara & Køien 2015).

Salakuuntelulla (eavesdropping) pyritään kuuntelemaan keskusteluja kahden osapuolen välillä. Seurannalla (tracking) käyttäjän liikkeitä voidaan jäljittää hänen laitteensa uniikilla identiteettinumerolla (unique identification number, UID). Seuraamalla käyttäjän sijaintia voidaan tunnistaa kohde tilanteissa, joissa käyttäjä haluaa pysyä anonyyminä. Salasanoihin perustuvissa hyökkäyksissä tunkeutujat yrittävät monistaa käyttäjän salasanan. Tämä voidaan tehdä kahdella tavalla, sanakirjahyökkäyksellä, jossa pyritään kokeilemaan mahdollisimman monta eri yhdistelmää numeroita ja kirjaimia, jotta voidaan arvata käyttäjän salasana, tai väsytyshyökkäyksellä, jossa käytetään hakkerointityökaluja, joilla

voidaan paljastaa kaikilla mahdollisilla salasanojen yhdistelmillä voimassa olevat salasanat (Abomhara & Køien 2015).

Tahsienin ym. (2020) mukaan aktiivisiin hyökkäyksiin kuuluvat muun muassa palvelunestohyökkäys, mies välissä -hyökkäys (man-in-the-middle attack), sybil -hyökkäys, spoofing -hyökkäys, watering hole -hyökkäys, verkonhäirintä (jamming), selective forwarding -hyökkäys, haitalliset syötteet (malicious inputs) sekä tietojen peukalointi. Spoofing ja sybil -hyökkäykset kohdistetaan käyttäjien tunnistetietoja, kuten RFID:tä ja MAC-osoitetta kohtaan, jotta saadaan luvaton pääsy IoT-järjestelmään. TCP/IP- yhteyskäytänteet eivät sisällä vahvoja suojausprotokollia, jotka tekevät IoT-laitteista erityisen haavoittuvaisia varsinkin spoofing -hyökkäyksille. Lisäksi sekä spoofing että sybil -hyökkäyksien avulla voidaan käynnistää lisähyökkäyksiä, kuten palvelunestohyökkäyksiä ja mies välissä -hyökkäyksiä (Tahsien ym., 2020). Mies välissä -hyökkäyksessä hyökätään kahden osapuolen väliseen kommunikaatioon asettumalla näiden väliin (Jansson & Sihvonen, 2018).

4 TEKOÄLYN KÄYTTÖ ESINEIDEN INTERNETIN KYBERTURVALLISUUDEN PARANTAMISESSA

Tässä luvussa käydään läpi tekoälyn ratkaisuja koneoppimisen avulla esineiden internetin laitteiden aiheuttamiin erityisiin kyberturvaongelmiin. Luku 4.1 perehtyy koneoppimisen soveltamisen mahdollisuuksiin esineiden internetin kyberturvallisuudessa, luku 4.2 koneoppimisen soveltamisen haasteisiin IoT-toimintaympäristössä, ja luku 4.3 koneoppimisen tulevaisuuden mahdollisuuksiin esineiden internetin kyberturvallisuudessa.

4.1 Koneoppimisen soveltaminen esineiden internetin kyberturvallisuudessa

Esineiden internet on hyvin erityinen sen ominaisuuksien puolesta verrattuna perinteiseen internetiin ja sen kyberturvallisuusratkaisuihin. Boukerche ja Coutinho (2021) esittävät, että kriittinen tosiasia on se, että perinteiset kyberturvallisuuden lähestymistavat eivät välttämättä ole tarpeeksi sopivia esineiden internetin sovelluksiin, sillä esineiden internetin laitteiden ja verkkojen ominaisuudet ovat ainutlaatuisia. IoT-laitteet ovat yhteydessä laajamittaisesti toisiin laitteisiin ja muihin rajapintoihin. Tämän yhteenliittymän takia ne vaativat matalatehoisia ja edullisia ratkaisuja, joten monimutkaisia turvallisuusmekanismeja ei voida käyttää esineiden internetin kyberturvallisuuden takaamiseen (Wu ym., 2020). IoT-verkot tuottavat valtavan määrän dataa, jota koneoppimisen ja syväoppimisen lähestymistavat vaativat, jotta ne voivat tuottaa älykkyyttä IoT-järjestelmiin. Koneoppimista käytetään suuressa määrin kyberturvallisuuteen, yksityisyyteen, kyberhyökkäysten tunnistukseen ja haittaohjelmien analysointiin (Hussain ym., 2020). Koneoppimisen menetelmiä, kuten ohjattua oppimista, ohjaamatonta oppimista ja vahvistusoppimista on laajalti käytetty parantamaan tietoverkkoturvallisuutta, esimerkiksi haittaohjelmien tunnistuksessa, häirinnänestossa, todentamisessa ja käyttöoikeuksien hallinnassa (Xiao ym., 2018).

Koneoppimisella on suunnatonta potentiaalia olla pääteknologia esineiden internetissä, sillä se tarjoaa arvokasta analytiikkaa tukemaan IoT-sovelluksia (Cui ym., 2018). Koneoppiminen voi auttaa koneita ja älylaitteita erittelemään hyödyllistä tietoa laitteiden tuottamasta tai ihmisten tuottamasta datasta. Koneoppimisen menetelmiä on käytetty muun muassa luokitteluun, regressioon ja tiheyden estimointiin (Hussain ym., 2020). Koneoppimisen metodit, kuten ohjattu oppiminen ja ohjaamaton oppiminen tarjoavat tehokkaan kyvyn tunnistaa epänormaalia toimintaa ja erottaa epätavallisia kaavoja IoT-verkossa, jonka avulla voidaan luokitella normaali käyttäytyminen ja epänormaalit hyökkäykset. Ohjattua ja ohjaamatonta oppimista voidaan siis käyttää laajasti esineiden internetin kyberturvallisuudessa (Wu ym., 2020). Koneoppiminen voi toimia analyyttisessa

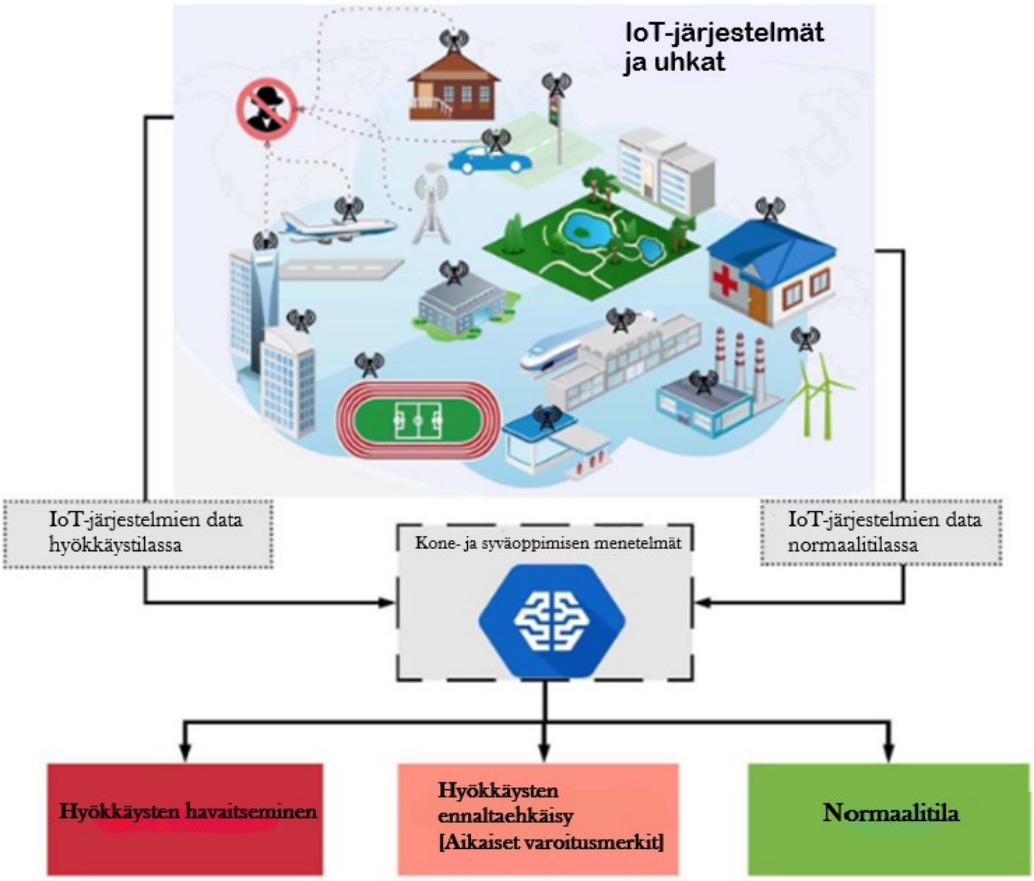
roolissa älykkäiden IoT-järjestelmien rakennuksessa, jotta voidaan tuottaa älykkäitä palveluita esineiden internetin maailmassa. Toisin kuin perinteiset analyyttiset metodit, koneoppiminen voi tehokkaasti selvittää huomaamatta jääneitä havaintoja Big datasta ja muuntaa Big dataa hyödylliseksi dataksi minimaalisella ihmisavustuksella (Al-Garadi ym., 2020). Big data on massadataa, joka koostuu jäsentämättömästä ja jäsennetystä, monimuotoisesta datasta ja on laajuudeltaan valtavaa. Big dataa eivät pysty käsittelemään perinteiset menetelmät datan käsittelyyn.

Koneoppiminen ja syväoppiminen ovat tehokkaita menetelmiä datan tutkiskelua varten, jonka avulla voidaan oppia IoT-komponenttien ja laitteiden normaalista ja epänormaalista käyttäytymisestä kun ne ovat vuorovaikutuksessa keskenään esineiden internetin toimintaympäristössä. Jokaisen IoT-järjestelmän osan syötedataa voidaan kerätä ja tutkia, jotta voidaan päätellä tavanomaisen vuorovaikutuksen kaavat, millä voidaan ennakoida ja tunnistaa haitallista käyttäytymistä varhaisessa vaiheessa. Lisäksi koneoppimisen ja syväoppimisen menetelmät saattavat olla tärkeitä uusien hyökkäysten tunnistamisessa, jotka ovat usein aikaisempien hyökkäysten muunnoksia, sillä ne voivat älykkäästi ennustaa uusia tuntemattomia hyökkäyksiä oppimalla olemassa olevista esimerkeistä (Al-Garadi ym., 2020). Kuluttajien IoT-laitteet, kuten älykkäät kodinkoneet ja puettavat laitteet, ovat vaarassa joutua hajautettujen palvelunestohyökkäysien kohteeksi. Koneoppiminen voi havainnoida kuluttajien IoT-laitteiden verkkoliikennettä, josta se voi erottaa sellaisen verkkoliikenteen, esimerkiksi hajautetuiden palvelunestohyökkäysten verkkoliikenteen, joka eroaa huomattavasti tavallisista verkkoliikenteen piirteistä, jota esiintyy kuluttajien IoT-laitteissa, ja käyttää tätä tietoa verkkoliikenteen erovaisuuksista hyökkäysten havaitsemiseen (Wu ym. 2020). IoT-laitteissa voidaan soveltaa ohjatun oppimisen menetelmiä, joilla voidaan arvioida sovellusten ajoaikaista käyttäytymistä haittaohjelmien havaitsemiseksi (Xiao ym., 2018).

Koneoppimisen menetelmillä on mahdollista havaita ja luokitella verkkotunkeutumisia ja hyökkäyksiä IoT-laitteita kohtaan. Xiaon ym. (2018) mukaan ohjatun oppimisen menetelmiä kuten tukivektorikoneita, naive Bayes algoritmeja, k-NN:ää, neuroverkkoja, syväoppimisen neuroverkkoja (deep Neural Network, DNN) ja satunnaismetsää voidaan käyttää tietoverkkoliikenteen tai IoT-laitteiden sovellusjälkien luokitteluun, jonka avulla voidaan rakentaa luokittelutai regressiomalleja. IoT-laitteet voivat käyttää esimerkiksi tukivektorikoneita havaitsemaan tietoverkkotunkeutumisia ja spoofing -hyökkäyksiä. IoT-laitteet voivat käyttää myös K-NN:ää verkkotunkeutumisen ja haittaohjelmien havaitsemiseen, neuroverkkoja havaitsemaan verkkotunkeutumisia ja palvelunestohyökkäyksiä. Naive Baeysia voidaan soveltaa esineiden internetin laitteissa havaitsemaan tunkeutumisia ja satunnaismetsän luokitinta voidaan käyttää haittaohjelmien havaitsemiseen. IoT-laitteet, joilla on tarpeeksi laskentakykyä ja muistiresursseja, voivat käyttää syviä neuroverkkoja havaitsemaan spoofing -hyökkäyksiä (Xiao ym., 2018). IoT-laitteiden turvaaminen palvelunestohyökkäyksiltä käyttämällä syväoppimisen menetelmiä on tehokas lähestymistapa, joka tuottaa merkittäviä tuloksia. Myös tyypilliset koneoppimisen menetelmät ovat

tehokkaita, mutta joiden toteuttaminen on yleensä kallista ja joiden tarkkuus ei muutu saavutettaessa tietty kynnys, johtuen IoT-verkkojen tuottamasta valtavasta datamäärästä (Ahmad & Alsmadi, 2021). IoT-laitteilla, kuten ulkoilmasensoreilla, on yleensä niukasti resursseja ja laskentakykyrajoitteita, joka heikentää tunkeutumisien havaitseminen tehokkuutta IoT-järjestelmissä. Koneoppimisen menetelmillä avulla voidaan rakentaa kevyitä käyttöoikeuksien hallintaprotokollia säästämään virtaa ja pidentämään IoT-järjestelmien elinaikaa (Xiao ym., 2018).

Koneoppimista voidaan käyttää IoT-järjestelmissä myös todentamiseen. Koneoppimiseen pohjautuva kyberturvallisuus voi parantaa IoT-järjestelmään lisättyjen uusien laitteiden todennusmekanismeja ja käyttöoikeuksien hallintaa tietoverkkoihin sekä tietoihin. Koneoppimista on käytetty myös ehdotetuissa ratkaisuissa todentaa IoT-laitteet tunnisteiden, tai ”sormenjälkien” avulla. (Boukerche & Coutinho, 2021). Todentaminen on yksi tärkeimmistä turvallisuusvaatimuksista esineiden internetissä. Käyttäjien on oltava todennettuja käyttääkseen esineiden internetin palveluja ja sovelluksia. IoT-sovellukset ja palvelut perustuvat tyypillisesti eri alustojen väliseen datan vaihtoon. Kun käyttäjän tai sovelluksen tarvitsee pyytää dataa IoT-laitteelta, tämän tahon on oltava todennettu IoT-verkossa ja täytyy varmistaa, että datan pyytäjällä on vaaditut käyttöoikeudet tietoihin tavalla, jossa järjestelmän yleiskäyttöisyys ei kärsi. Muussa tapauksessa pyyntö dataan evätään. On tärkeää estää ja myöntää tietyille käyttäjille pääsy esineiden internetin palvelujen ja sovellusten kriittisiin tietoihin (Hussain ym., 2020). Tunnisteilla todentaminen eli sormenjäljentäminen (fingerprinting) tarkoittaa laitteiden konfiguraation tunnistamista joko tutkimalla aktiivisesti laitteita viesteillä tai passiivisesti tarkkailemalla niiden luonnollista käyttäytymistä. Kohdeverkossa järjestelmävalvojat voivat käyttää sormenjäljennystä havainnoimaan ja määrittämään ongelmia, mutta myös hyökkääjät selvittämään verkon haavoittuvuudet (Wolf & Serpanos, 2018).



Kuvio 3 Esimerkkikuvitus koneoppimisen ja syväoppimisen mahdollisesta roolista IoT-turvallisuudessa (suom. Al-Garadi ym., 2020, 153827)

4.2 Koneoppimisen soveltamisen haasteet esineiden internetissä

Koneoppimisen menetelmien soveltamiseen esineiden internetissä sisältyy myös haasteita. Useimmat perinteisistä koneoppimisen menetelmistä eivät ole itsessään tehokkaita ja tarpeeksi skaalautuvia hallinnoimaan esineiden internetin tuottamaa dataa ja tarvitsevat siten huomattavia muutoksia. IoT-laitteet ovat pieniä ja tyypillisesti niitä rajoittaa virrankulutus sekä prosessointikyky, tämän takia koneoppimisen menetelmien soveltaminen sellaisenaan ei toimi sellaisessa resurssirajoitteisessa ympäristössä, missä laitteet toimivat. IoT-verkkojen tuottama data on luonteeltaan monipuolista, heterogeenistä, sisältäen eri formaatteja, tietotyyppejä, tiedostomuotoja, merkistöjä, tietomalleja ja datan eri tulkintatapoja ja merkityksiä. Datan heterogeenisyys johtaa ongelmiin tehokkaan ja yhtenäisen yleistyksen kannalta, erityisesti Big datassa ja erilaisissa tietojoukoissa, joilla on erilaisia ominaisuuksia. Tällainen data vaatii esikäsittelyä ja siivousta. Lisäksi koneoppimisalgoritmin valitseminen tiettyyn tapaukseen voi olla vaikeaa, sillä

esineiden internetin tuottama data voi olla nimiöityä, puoliksi jäsenneltyä, jäsentämätöntä tai jäsennettyä (Hussain ym, 2020). Jäsenneltyä dataa ovat esimerkiksi perinteisten tietokantojen taulut sarakkeineen ja riveineen, ja puolijäsenneltyä dataa ovat esimerkiksi HTML ja XML-tiedostot. Jäsentämätöntä dataa ovat muun muassa kuvat ja videot (Wu ym. 2020). Esineiden internetin laitteiden monipuolisuus ja monimutkaisuus tekevät sen tuottamasta datasta heterogeenistä, joka tekee myös tietopolun määrittämisestä IoT-verkoissa haastavampaa kuin perinteisissä tietoverkoissa. Todellisuudessa IoT-verkoissa suurin osa datasta on yleensä jäsentämättömässä muodossa (Cui ym., 2018).

Algoritmin valinta tehdään datan perusteella ja esimerkiksi ohjattua oppimista voidaan käyttää nimiöityyn dataan. Ongelmia aiheuttaa myös puute korkeatehoisista ja virtatehokkaista moniytimisistä mikroprosessoreista, jotka ovat suunniteltu koneoppimisen ja syväoppimisen neuroverkkoihin, jolla saadaan hyödynnettyä koneoppimisen potentiaali (Hussain ym., 2020). Minkä tahansa edistyneen koneoppimisen algoritmin kokoaminen on haastavaa, sillä se vaatii suuren määrän muistia ja kuluttaa lisävirtaa laajojen IoT-järjestelmien käsittelyn aikana. IoT-laitteet käsittelevät suuria datajoukkoja rajoitetuilla resursseilla. Lisäksi jos koneoppimisen algoritmeja sisällytetään IoT-järjestelmään, ne aiheuttavat lisää laskennallista monimutkaisuutta järjestelmälle, minkä takia on tarve pyrkiä minisoimaan tämä monimutkaisuus helpottamalla sitä koneoppimisen menetelmillä (Tahsien ym., 2020).

Tekoälyn menetelmät, jota voidaan käyttää esineiden internetin turvallisuutta yllä, sisältävät myös eri asteisesti turvallisuusriskejä niin kuin esineiden internet itsessään. Esimerkiksi hyökkääjä saattaa huijata koneoppimisen algoritmia, jotta hyökkääjää ei havaita. Tämä riippuu koneoppimisalgoritmin heikkouksista. Useimmat syväoppimisen menetelmät vaativat korkealaatuista dataa, jotta ne toimivat tehokkaasti. Tämän datan löytäminen on tärkeä haaste sekä koneoppimiselle että syväoppimiselle. Lisäksi myös datan ylläpitäminen on haaste. Johtuen esineiden internetin nopeasta kyvystä luoda dataa sekä datalähteiden monimuotoisuuden takia, korkeatasoisen datan ylläpitäminen reaaliajassa on haaste. Uuden hankitun datan tulisi olla samalaatuista kuin alkuperäisen datan, muuten alkuperäisen harjoitusmallin piirreavaruus (feature space) tuhoutuu pikkuhiljaa, joka johtaa mallivikaan. Koneoppimisen mallit ovat lisäksi aina ainutlaatuisia niiden kohdesovellukseen. Jos tietyssä ympäristössä hyväksi todettu malli siirretään samankaltaisiin ongelmiin, mitä se pyrkii ratkaisemaan, sen alkuperäiset parametrit saattavat pettää. Sen takia on tärkeää kouluttaa uusia malleja korvaamaan alkuperäiset mallit. Koneoppimisessa ja syväoppimisessa lisäksi hankalaa on se, että pienetkin muutokset algoritmissa syötetiedoissa saattavat muuttaa tulostetta suuresti, jonka takia on tärkeää huolehtia syötetietojen eheydestä ja vakaudesta, joka voi olla haastavaa IoT-toimintaympäristön tuottaessa suuren määrän lyhyin aikavälein julkistettavaa dataa (Wu ym., 2020).

4.3 Tulevaisuuden näkymä koneoppimiselle esineiden internetin kyberturvallisuudelle

Haasteista huolimatta, koneoppimiselle on vankka tulevaisuus edessä, jos sen menetelmiä voidaan muokata esineiden internetin laitteille sopivammaksi. Boukerche ja Coutinho (2021) ehdottavat, että ratkaisuiden tulisi olla mahdollisimman kevyitä, sillä laitteilla on rajoitettu määrä resursseja laskentakyvyssä, tallennuskapasiteetissa ja virrankulutuksessa. Näiden ratkaisuiden tulisi lisäksi ottaa huomioon luotettavan datajoukon puute, jota käytetään koneoppimisen algoritmien kouluttamiseen ja kelpuutukseen. Sen lisäksi, voi olla tarpeen uudelleen kouluttaa ja päivittää koneoppimisperustaisen kyberturvallisuusratkaisun parametrit. Datanvaihto tällaisille ratkaisuille tulisi toteuttaa tavalla, joka ei ruuhkauta verkkoa (Boukerche & Coutinho, 2021).

Mitä rikkaampi datajoukko, jonka avulla kone- ja syväoppimisalgoritmit voivat oppia, sitä tarkempia niiden havainnot ja toiminta luonnollisesti ovat (Abomhara & Køien, 2015). Koneoppimiselle sopivista esineiden internetin datajoukoista on puute lähtevän ja tulevan tietoverkkoliikenteen, laitteiden toimintojen ja käyttäjien vuorovaikutusten suhteen. Lisäksi IoT-sovellusten hyökkäyksiin ja uhkiin liittyvistä datajoukoista on puute (Boukerche & Coutinho, 2021). IoT-toimintaympäristön ainutlaatuiset ominaisuudet edellyttävät huolellista koulutusta sekä mallinrakennusta tunkeutumisien havaitsemiseen. On siis tärkeää kouluttaa kone- ja syväoppimismenetelmiä esineiden internetille spesifisellä sekä erityyppistä hyökkäysliikennettä sisältävällä datajoukolla. Menetelmiä koulutetaan tyypillisesti vertailuarvodatajoukoilla (benchmark dataset) tai laboratorioympäristössä simuloidussa tietoverkkoliikenteessä. Vertailuarvodatajoukot eivät yleensä ole ajantasaisia uusien hyökkäystyyppien suhteen. Muun muassa esineiden internetin tietoliikennettä sisältävistä sekä ajantasaisista datajoukoista on puute (Ahmad & Alsmadi, 2021).

Jotta saataisiin riittävän hyvin sovellettua koneoppimisen algoritmeja IoT-järjestelmiin, tarvitaan riittävät datajoukot, joita on vaikea kerätä sen perusteella, pystyykö järjestelmä tunnistamaan uhkia ja ryhtymään tarvittaviin toimiin. Tässä suhteessa data-augmentointi on yksi mahdollinen menetelmä ratkaisemaan datajoukkojen puutetta, jotta saadaan luotua tarpeeksi kattava datajoukko oikeaan dataan perustuen (Tahsien ym., 2020). Abomhara ja Køien (2015) ehdottavat, että yksi tapa millä voitaisiin luoda kattavampia datajoukkoja esineiden internetin uhkista ja hyökkäyksistä on joukkoistaminen. Kone- sekä syväoppimisalgoritmien kouluttamista varten tulisi luoda monipuolisia datajoukkoja, jotka sisältävät lähes kaikki hyökkäysmallit. Joukkoistaminen voisi tarjota mahdollisuuden kehittää ja luoda vastaavanlaisia datajoukkoja, jotka voisivat toimia myös mittapuuna (benchmarking) uusien algoritmien tarkkuuden testaamiseen vertailtaessa uusia algoritmeja jo olemassa oleviin hyökkäysten havaitsemismenetelmiin (Abomhara Køien, 2015).

Yksi mahdollinen tulevaisuuden ratkaisu on Edge AI sirut, kuten Google Coral Edge TPU, joita tällä hetkellä kehittävät muun muassa Qualcomm,

NVIDIA, Google ja Huawei (Wu ym., 2020). Yritysjohtamisen konsulttiyhtiön Advianin blogin mukaan Edge AI tarkoittaa reunalaskennan tai Edge-laskennan ja tekoälyn yhtymää (Kukkonen, 2020). Useimmat perinteisistä tekoälyn laskentatehtävistä suoritetaan etänä keskitetyillä ydinlaitteilla tai alustoilla, mutta tämä lähestymistapa ei ole optimaalisin ratkaisu esineiden internetille. Edge AI sirujen avulla pystytään upottamaan tekoälylaskentaa IoT-laitteisiin, sillä sirut mahdollistavat päätelaitteiden kyvyn suorittaa tekoälylaskentaa paikallisesti, joka vähentää siirtokuluja. Edge AI sirut voivat myös vähentää viivettä. Edge-laskenta vaatii reaaliaikaista toimintaa, sillä useiden eri laitteiden tulee tehdä reaaliaikaisia päätöksiä (Wu ym., 2020). Edge laskenta voisi mahdollistaa IoT-sovellusten asettamat vaatimukset verkon suojaukselle sekä latenssille ja kaistanleveydelle IoT-verkossa (Cui ym., 2018).

Tämänhetkiset koneoppimispohjaiset kyberturvallisuusratkaisut esineiden internetin todentamiseen ja pääsyoikeuksien hallintaan arvioivat laiteprofiileja niiden laitteiston epätäydellisyyksien perusteella. Laitteiden profilointiprosessiin voisi harkita lisätiedon hankkimista parempaa laiteprofilointia varten parantamaan esineiden internetin kyberturvallisuusratkaisuiden suorituskykyä. Esimerkiksi IoT-infrastruktuurin käyttötietoja, kuten prosessoreiden ja muistin resurssienkäyttöä ja tietoverkkoliikenteen voimakkuutta ja kaavoja sen sijaan, että keskistytään yhteen näkökohtaan, kuten nykyisessä lähdekirjallisuudessa. Myös korkeatasoista tietoa, kuten sosiaalista vuorovaikutusta toisten laitteiden kanssa, voitaisiin käyttää parantamaan koneoppimiseen perustuvien kyberturvallisuusratkaisuiden tehokkuutta sitä sovellettaessa esineiden internetiin (Boukerche & Coutinho, 2021).

Boukerche ja Coutinho (2021) ehdottavat että yhteistyöhön perustuvia sekä hajautettuja koneoppimispohjaisia kyberturvallisuusratkaisuja ei ole tutkittu. Esineiden internet tulee vaatimaan koneoppimispohjaisia ratkaisuja heterogeenisille ja hajautetuille laitteille. Tällaisten ratkaisujen on oltava yhteistoiminnallisia ja ne eivät voi perustua keskitettyyn koulutukseen datan avulla. Tässä mielessä yhdistettyä oppimista (federated learning) voitaisiin käyttää lähtökohtana näihin lähestymistapoihin. Sen lisäksi on kohdattava klassiset koneoppimisen haasteet, kuten puute datajoukoista koulutusta sekä validointia varten (Boukerche & Coutinho, 2021). Yhdeksi tulevaisuuden tutkimussuunnaksi ehdottavat Ahmad ja Alsmadi (2021) erilaisten kone- sekä syväoppimismenetelmien yhdistelyä optimaalisen hyökkäysten havainnoinnin sekä tehokkuuden saavuttamiseksi.

Uprety ja Rawat (2020) mukaan vahvistusoppiminen on ainoa koneoppimiseen kuuluvista menetelmistä, joka voi oppia ilman aikaisempia datajoukkoja, joita tyypillisesti käytetään koneoppisalgoritmien kouluttamiseen. Vahvistusoppiminen voisi siten vastata datajoukkojen puutteellisuuden ongelmaan. Esineiden internetin toimintaympäristö on monimutkainen ja siten vaikea mallintaa. Vahvistusoppiminen pystyy minimoimaan tällaisen toimintaympäristön mallinnukseen liittyvän vaivan, sillä vahvistusoppimisalgoritmi oppii toimintaympäristön mallin yrityksen ja erehdyksen avulla. Tämä antaa edun vahvistusoppimiselle, sillä ohjatun ja ohjaamattoman oppimisen algoritmit tarvitsevat koulutusta

datajoukon avulla, joka pitää ensin tuottaa algoritmeille. Tiedonkeruu on myös erityisen hankalaa joissakin esineiden internetin toimintaympäristöissä. Näissä tapauksissa ei ole olemassa valmista datajoukkoa, jolla kouluttaa muuhun kuin vahvistusoppimiseen kuuluva koneoppimisalgoritmi. Eritoten syvä vahvistusoppiminen voisi mallintaa esineiden internetin moniosaista toimintaympäristöä (Uprety & Rawat, 2020).

5 YHTEENVETO

Tutkielman tavoitteena oli muodostaa kokonaiskuva koneoppimisen soveltamisesta esineiden internetin kyberturvallisuuteen tällä hetkellä ja sen tulevaisuuden mahdollisuuksista vastaamalla tutkimuskysymyksiin. Tutkielman tutkimuskysymykset, joihin tutkielmassa pyrittiin vastaamaan, olivat *"Millä keinoilla koneoppiminen voi parantaa ja kehittää esineiden internetin kyberturvallisuutta?" ja "Mitä mahdollisuuksia koneoppimisella on tulevaisuudessa esineiden internetin kyberturvallisuuden parantamiseen?"*. Nykyinen kirjallisuus koneoppimisen ja syväoppimisen käytöstä esineiden internetin kyberturvallisuudessa kattaa IoT:n turvallisuuden tutkimalla olemassa olevia perinteisiä kyberturvallisuusratkaisuja ja uusien nousevien teknologioiden tarjoamia ratkaisuja. Perinteiset turvallisuus- ja yksityisyysratkaisut kärsivät kuitenkin useista esineiden internetin verkkojen dynaamiseen luonteeseen liittyvistä ongelmista. Vaikka IoT-verkoissa sovelletuista kone- ja syväoppimiseen pohjautuvista menetelmistä on runsaasti saatavilla olevaa kirjallisuutta, harvat tutkimukset kattavat koneoppimiseen ja syväoppimiseen perustuvia ratkaisuja. Lisäksi vaikka koneoppimista ja syväoppimista onkin käsitelty tutkimuksissa, niiden laajasta käytöstä esineiden internetin turvallisuudessa on yleisesti vähän tietoa (Hussain ym., 2020). Lähdekirjallisuudesta nousi esiin useita viittauksia ja kirjallisuuskatsauksen muodossa olevia kokoelmia aikaisemmin tehdyistä tutkimuksista, jotka kokeilivat koneoppimisen erilaisia menetelmiä IoT-järjestelmien turvaamiseksi tunkeilijoilta. Useista eri tutkimuksista ja eri menetelmistä huolimatta, ne eivät muodostaneet yhtä kattavaa standardia, millä puuttua esineiden internetin kyberhyökkäyksiin, vaan ne kokeilivat useita erilaisia metodeja. Tahsien ym. (2020) havainnoivat, että päätöspuut olivat eniten käytetty koneoppimisen menetelmä eri tutkimusten kokeilemissa IoT-kyberturvallisuusratkaisuissa verrattuna muihin menetelmiin. Päätöspuita eivät kuitenkaan käyttäneet enempää kuin 32 prosenttia kaikista tutkimuksista (Tahsien ym., 2020). Hussain ym. (2020) mukaan nykyiset tutkimukset, jotka keskittyvät koneoppimisen menetelmiin, ovat joko sovelluskohtaisia tai eivät kata IoT-verkkojen turvallisuuden ja yksityisyyden täyttä kirjoa.

Järjestelmät, jotka yhdistävät fyysisen maailman ja tietokonekomponentit luodakseen monimutkaisia käyttäytymismalleja kasvavat entistä enemmän

monimutkaisimmiksi ja kaikkialla läsnä oleviksi. Oletamme kuitenkin näiden järjestelmien tyydyttävän sekä suoja- että turvallisuusominaisuudet vaikka liian usein olemassa olevat järjestelmät eivät toteutakaan näitä vaatimuksia (Wolf & Serpanos, 2018). Esineiden internetin kohtaamat kyberturvallisuusuhkat muuttuvat ja kasvavat joka päivä IoT-toimintaympäristön muokkautuessa entistä enemmän ubiikiksi ja kytkeytyessä ihmisten, organisaatioiden ja valtioiden arkielämään. Tekoälypohjaisiin kyberturvallisuusratkaisuihin liittyviin haasteisiin on puututtava niitä suunnitellessa, jotta voidaan taata IoT-laitteiden turvallisuus (Boukerche & Coutinho, 2021).

Jatkotutkimuskohteiksi esineiden internetin kyberturvallisuudessa avoimia haasteita tunnistavat Sadique ym. (2018) olevan IoT-päätelaitteiden identiteetit asianmukaista todentamista ja valtuutusta varten, luottamuksen esineiden internetin paradigmassa eri komponenttien välillä, IoT-laitteiden tuottaman käyttäjädatan yksityisyys sekä esineiden internetin päästä päähän tietoturvallisuuden varmistamisen asianmukaisella turvallisuuden toimeenpanolla ja standardisoinnilla. Esineiden internetin kyberturvallisuusratkaisuiden tulee myös ottaa huomioon IoT-laitteiden resurssit, kuten virrankulutus, tallennuskapasiteetti ja laskentakyky. Reaaliaikaisten suojelu- ja havaitsemisjärjestelmien kehittäminen on tärkeää, jotta saadaan luotua tehokkaita turvallisuusmekanismeja, varsinkin suuriskaalaisille IoT-järjestelmille. Sen takia laskennallisen monimutkaisuuden vähentämisellä on tärkeä käytännön merkitys tulevalle tutkimukselle (Al-Garadi ym., 2020). Reaaliaikaisten kyberturvallisuusratkaisuiden tulisi olla sopivia IoT-laitteiden pienille resursseille.

Standardisointi kyberturvallisuusratkaisuille esineiden internetissä on puutteellinen, johtuen monista syistä, kuten esineiden internetin dynaamisesta luonteesta ja laitteiden heterogeenisyydestä. Myös datajoukkoja kone- ja syväoppimisen menetelmille on vielä niukasti saatavilla, tehden kone- ja syväoppimiseen perustuvien turvallisuusratkaisuiden tehokkuuden mittaamisesta vertailuanalyysin (benchmark) avulla vaikeaa (Hussain ym., 2020). Jatkotutkimukselle tärkeitä ja oleellisia aiheita ovatkin esineiden internetin koneoppimisen menetelmiin perustuvien kyberturvallisuusratkaisuiden standardointi, laskennallisen vaativuuden ja resurssien kulutuksen vähentäminen esineiden internetin koneoppimisen menetelmiin perustuvissa kyberturvallisuusratkaisuissa, sekä joukkoistamisen käyttö esineiden internetin uhkiin ja hyökkäyksiin liittyvien datajoukkojen tuottamiseksi, joilla voidaan kouluttaa koneoppimisen algoritmeja. Syväoppiminen ja vahvistusoppiminen nousivat lisäksi merkittäviksi menetelmiksi, joita tulisi tutkia suojausalgoritmeina ratkaisemaan esineiden internetin puutteellinen kyberturvallisuus. Kaiken lisäksi Edge AI on lupaava teknologia, jonka avulla voitaisiin minimoida koneoppimisen soveltamisen haasteita esineiden internetissä, kuten IoT-laitteiden resurssien rajallisuus.

LÄHTEET

Abomhara, M. & Køien, G. M. (2015). Cyber Security and the Internet of Things: Vulnerabilities, Threats and Attacks. *Journal of Cyber Security and Mobility*, 4, 65-88.

Ahmad, R., & Alsmadi, I. (2021). Machine learning approaches to IoT security: A systematic literature review. *Internet of Things*, 14, 2542-6605.

Ahmed, A. W., Ahmed, M. M., Khan, O. A. & Shah, M. A. (2017). A Comprehensive Analysis on the Security Threats and their Countermeasures of IoT. *International Journal of Advanced Computer Science and Applications*, 8(7), 489-501.

Alpaydin, E. (2014). *Introduction to Machine Learning* (2. uud. painos). London, England: The MIT Press.

Al-Garadi, M. A., Mohamed, A., Al-Ali, A. K., Du, X., Ali, I. & Guizani, M. (2020). A Survey of Machine and Deep Learning Methods for Internet of Things (IoT) Security. *IEEE Communications Surveys & Tutorials,* 22(3), 1646-1685.

Al Ghadeer, H. (2018). Cybersecurity Issues in Internet of Things and Countermeasures. *2018 IEEE International Conference on Industrial Internet (ICII)*, 195-201.

Balte, A., Kashid, A., & Patil, B. (2015). Security issues in Internet of things (IoT): A survey. *International Journal of Advanced Research in Computer Science and Software Engineering*, 5(4), 1-6.

Boukerche, A. & Coutinho, R. W. L. (2021). Design Guidelines for Machine Learning-based Cybersecurity in Internet of Things. *IEEE Network*, 35(1), 393-399.

Butun, I., Österberg, P. & Song, H. (2018) Security of the Internet of Things: Vulnerabilities, Attacks, and Countermeasures. *IEEE Communications Surveys & Tutorials*, 22(1), 616-644.

Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F. & Campbell, J. P. (2020). Introduction to Machine Learning, Neural Networks, and Deep Learning. *Translational vision science & technology*, 9(2):14.

Cui, L., Yang, S., Chen, F., Ming, Z., Lu, N., & Qin, J. (2018). A survey on application of machine learning for Internet of Things. *International Journal of Machine Learning and Cybernetics*, 9, 1399-1417.

Da Costa, K. A.P., Papa, J. P., Lisboa, C. O., Munoz, R., De Albuquerque, V. H. C. (2019). Internet of Things: A Survey on machine learning-based intrusion detection approaches. *Computer Networks*, 151, 147-157.

Empirica. (2020, 26. elokuuta). Mikä on IoT? Esineiden internet yksinkertaisesti selitettynä. Haettu 21.02.2023 osoitteesta https://www.empirica.fi/iot/

Hussain, F., Hussain, R., Syed, A. H., & Hossain, E. (2020). Machine learning in IoT security: Current solutions and future challenges. *IEEE Communications Surveys & Tutorials*, 22(3), 1686-1721.

International Telecommunications Union. (2008). *Overview of cybersecurity*. (ITU-T X.1205). Haettu osoitteesta https://www.itu.int/en/ITU-T/studygroups/com17/Pages/cybersecurity.aspx

International Telecommunications Union. (2012) *Overview of the Internet of Things*. (ITU-T Y.2060). Haettu osoitteesta https://www.itu.int/en/ITU-T/gsi/iot/Pages/default.aspx

Jansson, S., & Sihvonen, T. (2018). Kyberturvallisuus valtiollisena toimintaym päristönä ja siihen kohdistuvat uhkat. *Media & Viestintä*, *41*(1), 1-28.

Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. *Informatica Ljubljana*, 31(3), 249-268.

Kukkonen, S. (2020, 24. kesäkuuta). Advian blogi: Tekoälyä reunalla – Edge AI. Haettu 17.8.2021 osoitteesta https://www.advian.fi/blogi/edge-ai-mita-ja-miksi

Lu, Y. & Xu, L. D. (2019). Internet of Things (IoT) Cybersecurity Research: A Review of Current Research Topics. *IEEE Internet of Things Journal*, 6(2), 2103-2115.

Michalski, R., Carbonell, J. & Mitchell, T. (2013). *Machine Learning: An Artificial Intelligence Approach*. Springer Publishing Company.

Nazir, A., Sholla, S. & Bashir, A. (2019). Internet of Things Security: Issues, Challenges and Counter-Measures. *International Journal of Computing and Network Technology*, 7(3), 93-108.

Okoli, C., Schabram, K. (2010). A Guide to Conducting a Systematic Literature Review of Information Systems Research. *Sprouts: Working Papers on Information Systems*, 10(26).

Pantoja, M., Behrouzi, A. & Fabris, D. (2018). An Introduction to Deep Learning. *Concrete International: Farmington Hills*, 40(9), 35-41.

Portugal, I., Alencar, P. & Cowan, D. (2017). The Use of Machine Learning Algorithms in Recommender Systems: A Systematic Review. *Expert Systems with Applications*, 97, 205–227.

Puolustusministeriö (2013). *Suomen kyberturvallisuusstrategia*. Haettu osoitteesta https://turvallisuuskomitea.fi/wp-content/uploads/2018/05/Suomen-kyber-turvallisuusstrategia-ja-taustamuistio.pdf

Sadique, K. M., Rahmani, R. & Johannesson, P. (2018). Towards Security on Internet of Things: Applications and Challenges in Technology. *Procedia Computer Science*, 141, 199-206.

Salim, M. M., Shailendra, R. & Park, J. H. (2020). Distributed denial of service attacks and its defenses in IoT: A survey. *The Journal of Supercomputing, 76*(7), 5320-5363.

Samie, F., Bauer, L. & Henkel, J. (2019). From Cloud Down to Things: An Overview of Machine Learning in Internet of Things. *IEEE Internet of Things Journal*, 6(3), 4921-4934.

Samuel, A. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*.

Sessions, J. (2014). *The legal aspects of streaming digital media from the internet*. Unpublished Master's Thesis. Utica College.

Symantec Corporation. (2019). ISTR Internet Security Threat Report, 24. Haettu osoitteesta https://docs.broadcom.com/doc/istr-24-2019-en.

Tahsien, S. M., Karimipour, H. & Spachos, P. (2020). Machine learning based solutions for security of Internet of Things (IoT): A survey. *Journal of Network and Computer Applications*, 161.

The White House (2008). *National Security Presidential Directive 54/Homeland Security Presidential Directive 23* (NSPD54/HSPD23). Haettu osoitteesta https://fas.org/irp/offdocs/nspd/nspd-54.pdf

Uprety, A & Rawat, D. B. (2021). Reinforcement Learning for IoT security: A Comprehensive Survey. *IEEE Internet of Things Journal*, 8(11), 8693-8706.

Von Solms, R. & Van Niekerk, J. (2013). From information security to cyber security. *Computers and Security*, 38, 97-102.

Weber, R. H. & Studer, E. (2016). Cybersecurity in the Internet of Things: Legal aspects. *Computer Law & Security Review*, 35(5), 715-728.

Wolf, M. & Serpanos, D. (2018). Safety and Security in Cyber–Physical Systems and Internet-of-Things Systems. *Proceedings of the IEEE,* 106(1), 9-20.

Wu, H., Han, H., Wang, X. & Sun, S. (2020). Research on Artificial Intelligence Enhancing Internet of Things Security: A survey. *IEEE Access*, 8, 153826-153848.

Xiao, L., Wan, X., Lu, X., Zhang, Y. & Wu, D. (2018). IoT Security Techniques Based on Machine Learning: How Do IoT Devices Use AI to Enhance Security? *IEEE Signal Processing Magazine*, 35(5), 41-49.

Xin, Y., Kong, L., Liu, Z., Chen, Y. Li, Y., Zhu, H., Gao, M., Hou, M. & Wang, C. (2018). Machine Learning and Deep Learning Methods for Cybersecurity. *IEEE Access*, 6, 35365-35381.