Jonathan Kivimäki

Tekoäly poikkeavuuksien havaitsemisessa pilvi- ja tietokantaympäristöissä

Aikakausikatsaus koneoppimisen menetelmistä, aineistoista ja käyttöönoton haasteista

Tekniikan ja innovaatiojohtamisen akateeminen yksikkö
Kandidaatintutkielma
Automaatio ja tietotekniikka

Vaasa 2025

VAASAN YLIOPISTO
Tekniikan ja innovaatiojohtamisen akateeminen yksikkö
Tekijä: Jonathan Kivimäki
Tutkielman nimi: Tekoäly poikkeavuuksien havaitsemisessa pilvi- ja tietokantaympäristöissä: Aikakausikatsaus koneoppimisen menetelmistä, aineistoista ja käyttöönoton haasteista
Tutkinto: Tekniikan kandidaatti
Oppiaine: Automaatio ja tietotekniikka
Työn ohjaaja: Janne Koljonen
Valmistumisvuosi: 2025 **Sivumäärä:** 34

TIIVISTELMÄ:
Tutkielma tarkastelee tekoälypohjaista poikkeamien havaitsemista pilvi- ja tietokantaympäristöissä. Lähtökohtana on, että järjestelmien dynaamisuus, monivuokraus, heterogeeniset datalähteet ja salattu liikenne kaventavat sääntö- ja allekirjoituspohjaisten ratkaisujen kattavuutta. Tavoitteena on jäsentää keskeiset lähestymistavat, aineistot ja mittarit sekä arvioida mallien tuotantokelpoisuutta.

Tutkimus toteutetaan systemaattisena kirjallisuuskatsauksena vuosilta 2017–2025. Aineisto koostuu vertaisarvioiduista katsauksista ja soveltavista tutkimuksista, jotka käsittelevät pilvialustojen verkko- ja lokivirtoja, tietokantakyselyiden poikkeamia sekä tunkeutumisen havaitsemista. Menetelmällinen tarkastelu kattaa ohjatut ja ohjaamattomat mallit, syväoppimisen sekä niiden yhdistelmät. Arvioinnissa painotetaan epätasapainoisiin luokkajakaumiin sopivia mittareita, kuten havaitsemisherkkyys, väärien hälytysten osuus ja F1-lukua sekä operatiivisia vaatimuksia, kuten viivettä ja läpivirtauksen kestävyyttä, kustannuksia, selitettävyys ja tietosuoja.

Katsauksen mukaan ohjattu oppiminen toimii hyvin tunnettujen hyökkäysten havaitsemisessa, jos opetusdata kuvaa kohdeympäristön nykyistä liikennettä. Nollapäivä-hyökkäykset ja nopeasti muuntuvat ilmiöt puoltavat ohjaamatonta anomaliantunnistusta ja syväoppimista, mutta ne lisäävät tulkittavuus- ja hälytyshaasteita. Hybridit, joissa yhdistetään allekirjoitus- ja sääntöpohjainen tunnistus sekä koneoppimis- ja syväoppimispohjainen anomaliantunnistus, tasapainottavat havaitsemisherkkyyttä ja hälytyskuormaa. Tietokantatasolla tehokkaaksi osoittautuu kyselykäyttäytymisen profilointi; eheyden ja luottamuksellisuuden vuoksi matala väärien negatiivisten taso on kriittinen. Aineistoriippuvuus rajoittaa yleistettävyyttä, joten arviointi tulisi yhdistää julkisista dataseteistä saatuihin tuloksiin organisaatiokohtaisella tai realistisesti synteettisellä datalla. Tuotantokelpoisuus edellyttää selitettävyyttä ja tietosuojan varmistamista. Johtopäätöksenä suositellaan kerroksellista arkkitehtuuria, epätasapainoisiin luokkiin sopivia mittareita ja käyttöönottoa hallituissa dataputkissa. Jatkotutkimus tulee suunnata mallien siirrettävyyteen ja ympäristön ajallisten muutosten hallintaan. Erityisesti tietokantaympäristöissä jatkotutkimus tulisi kohdistaa kyselyprofiilien ja pääsykuvioiden muutosten havaitsemiseen sekä mallien siirrettävyyteen eri sovellusten välillä, vertaillen tuloksia sekä vakiintuneisiin aineistoihin että organisaatiokohtaisiin lokivirtoihin.

AVAINSANAT: kyberturvallisuus; koneoppiminen; syväoppiminen; anomaliat; tunnistaminen; verkkohyökkäykset; tietokannat; pilvipalvelut.

Sisällys

Kuvat

Taulukot

Lyhenteet

ACC - tarkkuus (Accuracy)
AE - autoenkooderi (Autoencoder)
AI - tekoäly (Artificial Intelligence)
API - sovellusohjelmointirajapinta (Application Programming Interface)
AUC - käyrän alle jäävä pinta-ala (Area Under the Curve)
CIA - luottamuksellisuus, eheys, saatavuus (Confidentiality, Integrity, Availability)
CNN - konvoluutioneuroverkko (Convolutional Neural Network)
CPU - Prosessori (Central Processing Unit)
DDoS - hajautettu palvelunestohyökkäys (Distributed Denial of Service)
DL - syväoppiminen (Deep Learning)
DoS - palvelunestohyökkäys (Denial of Service)
F1 - F1-luku (harmoninen keskiarvo tarkkuudesta ja kattavuudesta)
FN - väärä negatiivinen (False Negative)
FNR - väärien negatiivisten osuus (False Negative Rate)
FP - väärä positiivinen (False Positive)
FPR - väärien positiivisten osuus (False Positive Rate)
FTP - tiedonsiirtoprotokolla (File Transfer Protocol)
GAN - generatiivinen vastakkaiset verkot (Generative Adversarial Network)
GPU - näytönohjain (Graphics Processing Unit)
GRU - porttirakenteinen toistoyksikkö (Gated Recurrent Unit)
IaaS - infrastruktuuri palveluna (Infrastructure as a Service)
IDS - tunkeutumisen havaitsemisjärjestelmä (Intrusion Detection System)
IPS - tunkeutumisen estojärjestelmä (Intrusion Prevention System)
IoT - esineiden internet (Internet of Things)
LSTM - pitkä-lyhytkestoinen muisti (Long Short-Term Memory)
ML - koneoppiminen (Machine Learning)
NIDS - verkko-IDS (Network-based IDS)
PaaS - alusta palveluna (Platform as a Service)
PPV - positiivinen ennustearvo / täsmällisyys (Positive Predictive Value / Precision)
RAM - keskusmuisti (Random Access Memory)
RNN - toistuva neuroverkko (Recurrent Neural Network)
ROC - vastaanottajan toimintakäyrä (Receiver Operating Characteristic)
SaaS - ohjelmisto palveluna (Software as a Service)
SSH - Turvattu kuori (Secure Shell)
TP - tosi positiivinen (True Positive)
TPR - todellisten positiivisten osuus / kattavuus (True Positive Rate / Recall)
UEBA - käyttäjä- ja entiteettikäyttäytymisen analytiikka (User and Entity Behavior Analytics)
XAI - selitettävä tekoäly (Explainable AI)
XSS - sivustojen välinen skriptaus (Cross-Site Scripting)
LIME - Selitettävän tekoälyn metodi (Local Interpretable Model-agnostic Explanations)
SHAP - Selitettävän tekoälyn metodi (SHapley Additive exPlanations)
USD - Yhdysvaltain dollari

1 Johdanto

Tekoälyä (artificial intelligence, AI) ja etenkin koneoppimista (machine learning, ML) hyödynnetään yhä useammin kyberhyökkäyksissä, mikä kasvattaa hyökkäysten nopeutta, mittakaavaa ja muovautuvuutta (Aksela ja muut, 2022). Organisaatioiden siirtyessä laajasti pilvipalveluihin (cloud computing), herää kysymys, miten poikkeamat tunnistetaan jatkuvasti kehittyvässä ympäristössä. Ympäristössä, joissa dynaaminen skaalautuminen, monivuokraus ja heterogeeniset datalähteet muuttavat normaalia käyttäytymistä.

Tämä tutkielma vastaa kysymykseen vertaamalla AI-pohjaisia poikkeamien havaitsemismenetelmiä allekirjoitus- (signature-based) ja sääntöpohjaisiin (rule-based) tunkeutumisen havaitsemisratkaisuihin pilvi- ja tietokantaympäristöissä.

Sarker ja muut (2021) kuvaavat AI-ratkaisuja yhtenä neljännen teollisuuden vallankumouksen avainteknologiana. Tietoturvan peruspilarit - luottamuksellisuus, eheys ja saatavuus, eli CIA-luokittelu (Confidentiality, Integrity, and Availability) korostaa havaitsemisen nopeuden ja virhehälytysten hallinnan merkitystä pilvessä. Valtava määrä dataa syntyy ja kerätään yleistyvien teknologioiden, kuten esineiden internetin (Internet of Things, IoT) ja pilvipalveluiden yleistyessä. Kerättyä dataa voidaan hyödyntää yrityksen tai organisaation eduksi, mutta kyberhyökkäykset asettavat suuria haasteita. Kyberhyökkäys on tyypillisesti yhden henkilön tai organisaation pahantahtoinen ja koordinoitu yritys murtautua toisen henkilön tai organisaation tietojärjestelmään. Sarkerin (2021) esittelemä IBM-raportti arvioi keskimääräisen tietomurron kustannukseksi Yhdysvalloissa 8,19 miljoonaa USD ja kyberrikollisuuden globaalit vuosikustannukset ovat arviolta 400 miljardia USD, mikä korostaa skaalautuvien kyberturvallisuusratkaisujen tarvetta.

1.1 Tutkimuskysymykset sekä tavoitteet

Tämä tutkielma jäsentää AI-/ML-pohjaiset poikkeamien havaitsemisen lähestymistavat ja vertaa niitä allekirjoitus- ja sääntöpohjaisiin ratkaisuihin pilvi- ja tietokantaympäristöissä. Tutkimuksen pääkysymys on seuraava:

1. Kuinka tekoälypohjaiset poikkeamien havaitsemismallit voivat edistää tietoturvaa pilvi- ja tietokantaympäristöissä?

Tutkimuskysymykseen vastaa seuraavat alatutkimuskysymykset:

1. Mitä eroja tunnistettujen mallien välillä on tehokkuuden, tarkkuuden ja soveltuvuuden näkökulmasta pilvi- ja tietokantojen tietoturvaympäristössä
2. Mitä käytännön haasteita ja vaatimuksia liittyy tekoälypohjaisten poikkeamien havaitsemismallien käyttöönottoon pilvi- ja tietokantaympäristössä?

1.2 Menetelmät ja rakenne

Tämä tutkimus toteutetaan systemaattisena kirjallisuuskatsauksena. Aineisto tähän työhön on kerätty vuosien 2017–2025 välillä julkaistuista vertaisarvioiduista artikkeleista ja laajoista katsauksista, jotka keskittyvät kone- ja syväoppimisen käyttöön pilvi- ja tietokantaympäristöissä. Tiedonkeruu suoritettiin hyödyntäen IEEE Xplore, ACM Digital Library, Scopus ja SpringerLink -tietokannoista. Lisäksi hyödynnettiin lähdeluetteloiden läpikäyntiä ja Google Scholaria.

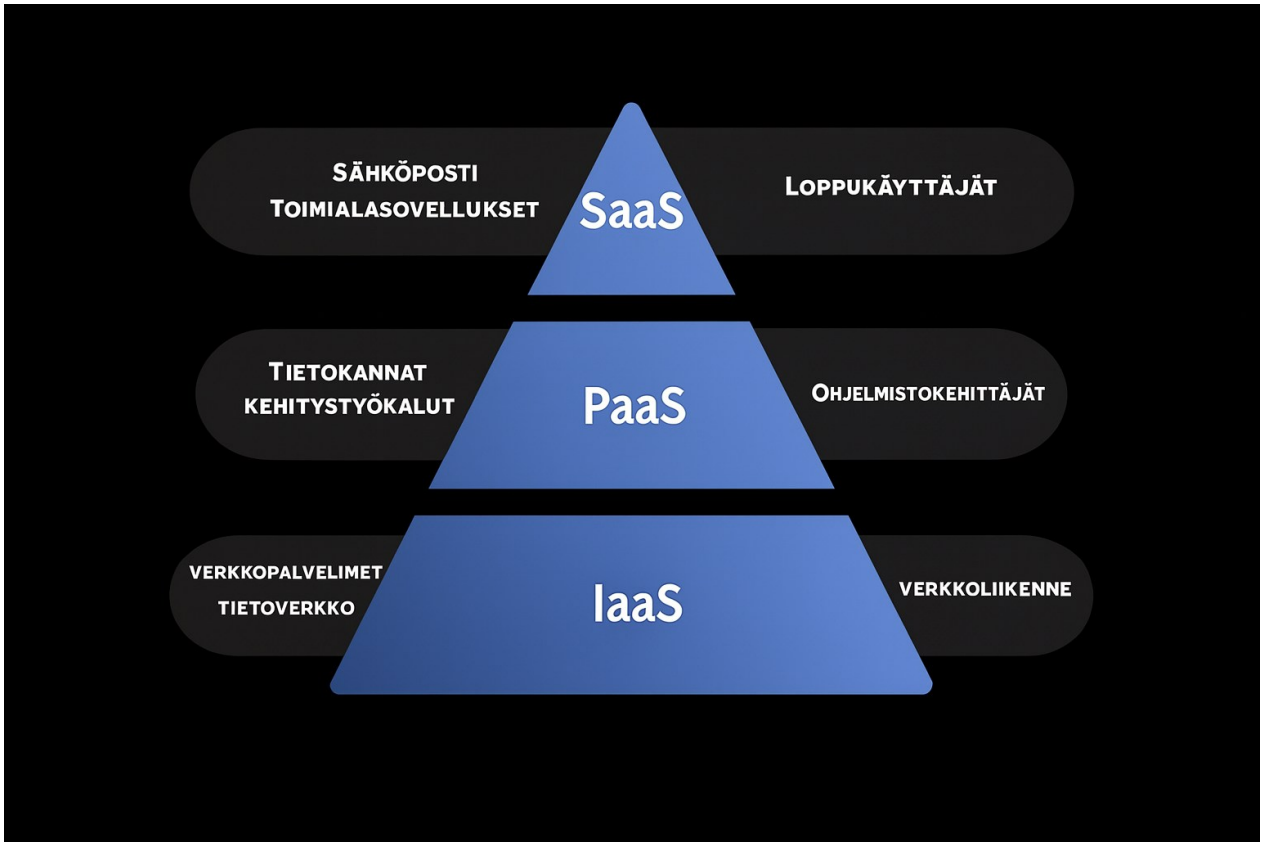
Tutkielman rakenne on seuraava: Luvussa 2 annetaan yleiskatsaus pilvipalveluiden arkkitehtuuriin ja tietoturvaympäristöön. Luvussa 3 esitellään erilaisia tilastolliset menetelmät sekä ML-mallit. Luvussa 4 arvioidaan mallien soveltuvuutta. Luvussa 5 tarkastellaan AI:n käyttöön liittyviä eettisiä ja lainsäädännöllisiä näkökulmia tietoturvassa. Luku 6 kokoaa tutkielman keskeiset tulokset ja esittää johtopäätökset.

2 Pilvipalveluiden tietoturvaympäristö ja turvauhkat

Pilvipalveluilla tarkoitetaan nykyaikaisia tietoteknisiä ratkaisuja, joiden avulla mm. laskentaresursseja, tallennustilaa ja ohjelmistopalveluita voidaan tarjota internetin välityksellä (Hu ja muut, 2017). Pilvipalvelumallit siirtävät havaintopisteitä verkosta kohti sovellusta, identiteettiä ja tietokantoja. Siksi poikkeamien havaitseminen on kytkettävä palvelumalliin ja jaettuun vastuuseen asiakkaan ja palveluntarjoajan välillä (Nassif ja muut 2021).

2.1 Pilvipalveluiden palvelumallit

Pilvipalvelut jakautuvat palvelumalleihin Software as a Service (SaaS), Platform as a Service (PaaS) ja Infrastructure as a Service (IaaS). Palvelumallit määritellään tarkemmin alaluvuissa 2.1.1–2.1.3. Nämä määrittävät, missä poikkeamien havaitsemisen (anomaly detection) kannalta keskeiset havaintopisteet sijaitsevat: verkossa, sovellus- ja identiteettitasolla sekä tietokantakyselyissä. Kuvassa 1 havainnollistetaan palvelumallit (Abdallah ja muut, 2024).



Kuva 1 Pilvipalvelumallit (Abdallah ja muut, 2024)

2.1.1 Ohjelmisto palveluna

Ohjelmisto palveluna-mallissa (Software as a Service, SaaS) ohjelmisto ja tietokanta sijaitsevat palveluntarjoajan hallinnoimassa ympäristössä ja palvelua käytetään tyypillisesti selaimella ilman asiakkaan omaa infrastruktuuria. Asiakkaan rajoiteussa infrastruktuurinäkymässä poikkeamien havaitseminen painottuu Käyttäjä- ja entiteettikäyttäytymisen analysointiin (User and Entity Behavior Analytics, UEBA). UEBA:n avulla voidaan seurata poikkeavia kirjautumismalleja, poikkeavia API-kutsujen telemetriaa (Application Programming Interface) tai (Abdallah ja muut, 2024). Monivuokraus lisää hajautettujen palvelunestohyökkäyksien (DDoS) riskiä ja siksi SaaS-ympäristöihin on ehdotettu monivuokraajille sovitettu IDS-ratkaisuja (Intrusion Detection System). (Mohamed, 2025).

SaaS-malli on laajalti käytössä ja siksi houkutteleva kohde uhkatoimijoille. SaaS-palvelut ovat keskitetysti isännöityjä ja niitä käytetään etäyhteyksin, täten ne ovat alttiita

palvelunestohyökkäyksille (DDoS). Laajoissa katsauksissa DDoS on pilviturvallisuuden yleisimmin tutkittujen ongelma-alueiden joukossa. Poikkeavuuksien havaitseminen ja varhainen reagointi ovat siksi keskeisessä osassa SaaS-ympäristöjen suojauksessa (Nassif ja muut, 2021; Abdallah ja muut, 2024).

2.1.2 Alusta palveluna

Alusta palveluna-malli (Platform as a Service, PaaS) tarjoaa kehitys- ja ajoympäristön, johon sisältyy tyypillisesti käyttöjärjestelmä, ohjelmointikielet, web-palvelimet ja tietokantapalvelut. Resurssit skaalautuvat automaattisesti tarpeiden mukaisesti ja sovelluksia kehitetään palveluntarjoajan API:n avulla käyttötarkoituksen mukaan. Tässä mallissa asiakas hallitsee ohjelmiston käyttöönottoa ja konfigurointia (Abdallah ja muut, 2024).

PaaS-mallin tietoturvahaasteet liittyvät erityisesti API-liikenteen profilointiin, sovelluskoodin haavoittuvuuksiin sekä tietokantojen suojaamiseen. Tunnusomaisia uhkia ovat API-väärinkäyttö, konfiguraatiovirheet sekä riippuvuuksien haavoittuvuudet. PaaS-ympäristössä on tutkittu ja toteutettu ML-pohjaista DDoS havaitsemista. (Abdallah ja muut, 2024).

2.1.3 Infrastruktuuri palveluna

Infrastruktuuri palveluna (Infrastructure as a Service, IaaS) tarjoaa käyttäjille virtuaalisia laskentaresursseja, kuten palvelimia, tallennustilaa ja verkkoyhteyksiä internetin välityksellä (Hu ja muut, 2017). Käyttäjä voi hallita ja ylläpitää näitä resursseja virtuaalisesti ilman tarvetta omistaa fyysistä infrastruktuuria. IaaS-palveluissa käyttäjällä on suuri vastuu omien sovellustensa, käyttöjärjestelmiensä sekä tietoturvan hallinnasta. IaaS-palveluiden tietoturvauhat liittyvät resurssien väärinkäyttöön, konfiguraatiovirheisiin sekä virtuaalisen infrastruktuurin haavoittuvuuksiin (Abdallah ja muut, 2024). DDoS-suojaus sekä verkon poikkeamien tunnistus ovat keskeisiä tuotantovalmiissa arkkitehtuureissa. ML-pohjaiset IDS-menetelmät täydentävät sääntö- ja allekirjoituspohjaisia ratkaisuja nopeasti muuttuvassa pilviympäristössä (Dong ja Kotenko, 2025).

2.2 Uhkamallit pilvi- ja tietokantaympäristöissä

Pilvipalvelut ovat monimutkaisia järjestelmiä, jotka altistuvat monenlaisille tietoturvariskeille. Järjestelmän monimutkaisuuden kasvaessa myös sen haavoittuvuudet lisääntyvät (Nassif ja muut, 2021). Keskeisiä pilviturvan teema-alueita ovat erityisesti palvelunesto-hyökkäykset sekä tietosuojan ja yksityisyyden ongelmat, jotka korostuvat datakeskeisissä työkuormissa. Nassifin (2021) mukaan DDoS ja datan yksityisyys ovat pilviturvan tutkituimpia osa-alueita, mikä kuvastaa niiden painoarvoa.

DDoS-uhkien mittakaavaa kuvaa Ciscon laajasti siteerattu ennuste, jossa DDoS-hyökkäysten määrän ennustettiin tuplaantuvan vuoteen 2023 mennessä, jolloin niitä olisi noin 15,3 miljoonaa vuodessa (Abdallah ja muut, 2024).

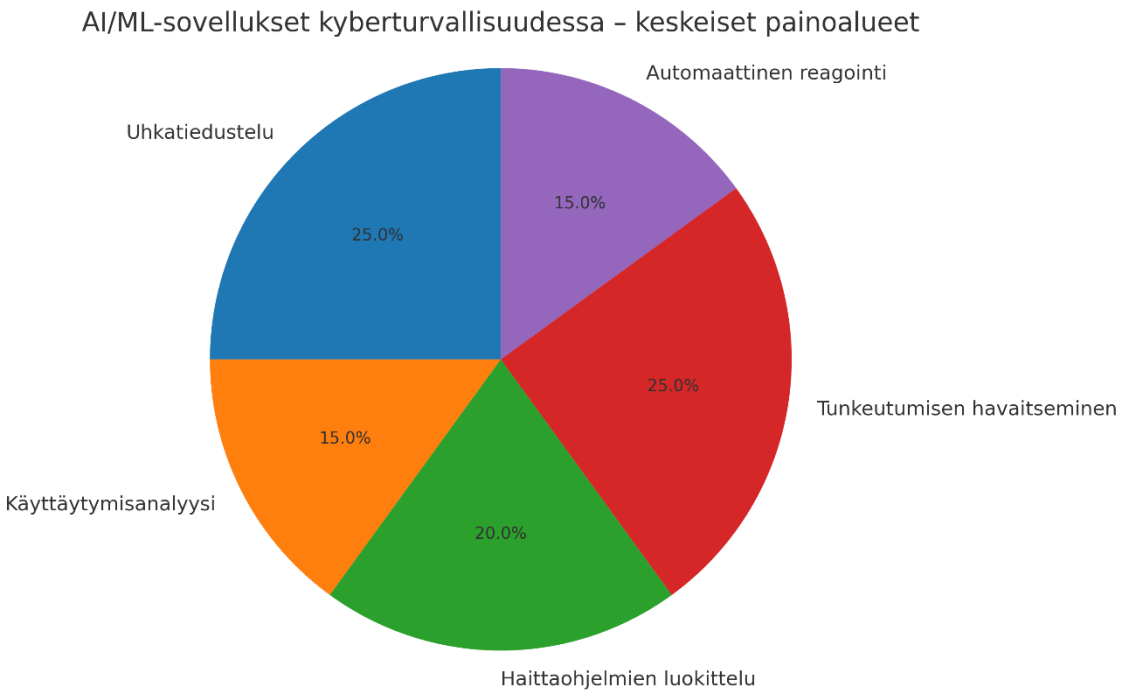
Tietokantaympäristöissä hyökkäykset kohdistuvat luottamuksellisuuteen ja eheyteen muun muassa SQL-injektion, liiallisten oikeuksien, epätavallisten kyselyprofiilien ja tiedon salakuljetuksen kautta. Sisäiset uhat ja väärät konfiguraatiot ovat toistuvia riskitekijöitä pilvessä, jossa monivuokraus ja dynaaminen skaalautuminen vaikeuttavat perinteistä valvontaa. Yleisessä uhkakuvassa esiintyvät myös phishing ja haittaohjelmat sekä identiteetin väärinkäyttö, jotka toimivat hyökkäysketjujen alkuvaiheina (Sarker ja muut 2021).

2.3 Poikkeamien havaitsemisen rooli tietoturvassa

Poikkeamien havaitseminen täydentää allekirjoitus- ja sääntöpohjaisia ratkaisuja tilanteissa, joissa hyökkäystapa muuttuu, kohde on uusi tai allekirjoitus puuttuu. Lähestymistapa rakentaa mallin normaalista käyttäytymisestä ja tunnistaa siitä poikkeavat tapahtumat verkossa, identiteetissä ja tietokantakyselyissä. Pilvessä tämä on keskeistä nollapäivähyökkäysten, API-väärinkäytön ja skaalaus- sekä konfiguraatiopoikkeamien havaitsemiseksi; tietokannoissa se paljastaa epätavanomaiset join-rakenteet, poikkeavat tulosjoukon koot ja aikataulukäytön muutokset. (Abdallah ja muut 2024).

3 Tekoälypohjaiset poikkeamien havaitsemismallit

AI tarjoaa tehokkaita menetelmiä pilvipalveluiden tietoturvapoikkeamien havaitsemiseen. Kyseiset menetelmät käyttävät laskentaa ja analytiikkaa tunnistamaan epätavallisia tapahtumia, joita perinteiset menetelmät kuten allekirjoitus tai sääntöpohjaiset menetelmät eivät välttämättä havaitsisi. (Ahmed ja muut, 2025). Kuvassa 2 on esitettynä AI/ML-sovelluksien keskeiset osa-alueet kyberturvassa. Näistä tämän tutkielman painopiste on tunkeutumisen havaitseminen ja käyttäytymisanalyysi, jotka kattavat noin 40 prosenttia osa-alueista.



Kuva 2 AI-/ML- keskeiset painoalueet tietoturvassa (Mohamed, 2025)

3.1 Perinteiset poikkeamien tunnistusmenetelmät

Poikkeamien havaitseminen tietoturvassa perustui pitkään perinteisiin menetelmiin, jotka jaotellaan vakiintuneesti kahteen luokkaan: allekirjoitus- tai väärinkäyttöpohjaisiin menetelmiin (signature-based / misuse detection), sekä anomaliapohjaiseen tunnistukseen (anomaly detection) (Halbouni ja muut, 2022). Nämä lähestymistavat muodostivat tunkeutumisen havaitsemisjärjestelmät (intrusion detection system, IDS) jo ennen tekoäly- ja koneoppimisratkaisujen yleistymistä.

Allekirjoituspohjaiset menetelmät vertaavat havaittua verkkoliikennettä ja järjestelmän tapahtumia tunnettuihin hyökkäysmalleihin eli allekirjoituksiin (Ahmed ja muut, 2025). Menetelmä on erittäin tarkka tunnettuja hyökkäyksiä vastaan ja tuottaa vähän vääriä hälytyksiä (false positives, FP). Sen keskeisiä rajoitteita ovat riippuvuus allekirjoitustietokannan jatkuvasta päivityksestä ja kyvyttömyys tunnistaa uusia, tuntemattomia hyökkäyksiä, kuten nollapäivähyökkäyksiä (zero-day attack), sillä näihin ei ole olemassa valmiita allekirjoituksia.

Anomaliapohjainen tunnistus rakentaa mallin normaalista käyttäytymisestä ja käyttää tätä havaintojen pohjana (Ahmed ja muut, 2025; Ahmetoglu ja Das, 2022). Kun käytös poikkeaa tästä mallista, se merkitään poikkeamaksi ja mahdollisesti haitalliseksi (Ahmetoglu ja Das, 2022). Anomaliapohjaisen tunnistuksen vahvuus on kyky tunnistaa ennennäkemättömiä uhkia, mutta käytännössä menetelmät kärsivät usein korkeasta FP määrästä, mikä heikentää käyttökelpoisuutta ja kuormittaa operatiivista työtä (Ahmed ja muut, 2025).

Perinteisten lähestymistapojen rakenteelliset rajoitteet ovat johtaneet hybridimalleihin, joissa tunnetut uhat katetaan allekirjoituksilla ja tuntemattomat uhat pyritään tavoittamaan anomaliapohjaisesti (Ahmetoglu ja Das, 2022). Samasta syystä kehitys on siirtynyt kohti AI- ja ML-ratkaisuja, joiden tavoitteena on parantaa kattavuutta ja vähentää väärien hälytyksien määrää dynaamisissa ympäristöissä (Halbouni ja muut, 2025).

3.2 Koneoppimismenetelmät

Koneoppiminen tai koneoppimismenetelmät ovat AI-ratkaisuja, jotka mahdollistavat tietokoneita oppimaan datan pohjalta ilman erillistä ohjelmointia (IBM, 2022). Tyypillisesti ML jaetaan kolmeen kategoriaan; ohjattu oppiminen (supervised learning), ohjaamaton oppiminen (unsupervised learning) ja vahvistusoppiminen (reinforcement learning) (IBM, 2022).

3.2.1 Ohjattu oppiminen

Ohjattu oppiminen perustuu tekoälyn kouluttamiseen merkityllä aineistolla, jossa normaalitilanteet sekä poikkeamat ovat ennalta tunnistettu. AI:lle näytetään esimerkkejä syötteistä (input) sekä halutuista tuloksista (output). Tavoitteena on kouluttaa tekoäly, joka kykenee päättelemään todennäköisen tuloksen tuntemattomien syötteiden avulla (Nasteski, 2017; Solin, 2022). Solinin (2022) mukaan ohjattua oppimista käytetään kouluttamaan regressio- ja luokittelumalleja. Regressiomalli pyrkii ennustamaan tai selittämään jatkuvia suureita, kuten esimerkiksi DDoS tai SQL-injectioiden tunnisteita. Jos koneoppimismallilla on ennalta määritelty määrä luokkia, kyseessä on luokittelumalli (Solin, 2022).

3.2.2 Ohjaamaton oppiminen

Ohjaamattomassa oppimisessa koneoppimismalli koulutetaan käyttämällä merkitsemätöntä tietoaineistoa, eli algoritmi analysoi tietoaineiston rakennetta löytääkseen piileviä kaavoja tai poikkeavaisuuksia (IBM, 2022; Mohamed, 2025). Mohamedin (2025) mukaan, ohjaamaton oppiminen on erityisen hyödyllinen poikkeavuuksien havaitsemisessa, jossa tarkoituksen on tunnistaa normaalista poikkeavaa toimintaa, ilman ennestään tunnettuja hyökkäyksen tunnisteita. Sarkerin (2021) mukaan ohjaamatonta oppimista käytetään lähinnä klusterointiin (clustering) sekä informatiivisiin datamuunnoksiin (informative data transformations).

Klusterointi on prosessi, jossa havaintoyksiköt ryhmitellään klustereiksi siten, että saman klusterin sisällä olevat havainnot ovat keskenään mahdollisimman samankaltaisia ja eri klustereissa mahdollisimman erilaisia. Klusterointi menetelmiä ovat esimerkiksi k-keskiarvot (k-means*)*, hierarkkinen klusterointi ja tiheysperustaiset menetelmät kuten DBSCAN. Pilvipalveluiden tietoturvassa klusterointia voidaan hyödyntää esimerkiksi verkon liikenteen, käyttäjäistuntojen tai palvelinlokien ryhmittelyssä, jolloin poikkeavat ryhmät voivat viitata epätyypilliseen käyttäytymiseen tai mahdollisiin hyökkäyksiin (Mohamed, 2025). Menetelmän vahvuutena on kyky löytää rakenteita ilman ennalta merkittyjä poikkeamia, mutta sen rajoitteisiin kuuluvat parametrien, kuten klusterien määrän tai etäisyysmitan valinnan herkkyys.

Informatiiviset datamuunnokset, kuten dimensioiden vähentäminen, pyrkivät tiivistämään monimutkaista ja moniulotteista dataa säilyttäen samalla sen olennaisimman informaatiosisällön. Tunnettuja menetelmiä ovat pääkomponenttianalyysi (Principal Component Analysis, PCA), joka säilyttää suurimman osan datan varianssista muutamassa pääkomponentissa, sekä epälineaariset menetelmät kuten t-jakautunut stokastinen naapurin upotus (t-distributed Stochastic Neighbor Embedding, t-SNE) ja Yhtenäinen monistoapproksimaatio ja projektio (Uniform Manifold Approximation and Projection, UMAP), jotka soveltuvat monimutkaisten rakenteiden visualisointiin ja ryhmien erottamiseen. Pilvipalveluiden tietoturvassa dimensioiden vähentäminen voi toimia esikäsittelyvaiheena ennen klusterointia, jolloin kohinaa voidaan poistaa ja poikkeamien havaitsemisen tarkkuutta parantaa (Mohamed, 2025). Menetelmän heikkoutena on mahdollinen informaation katoaminen ja tulosten tulkinnan vaikeus.

3.2.3 Vahvistusoppiminen

Vahvistusoppiminen (reinforcement learning) on koneoppimisen edistyneempi muoto. Malli oppii vuorovaikuttamalla ympäristöönsä, josta se vastaanottaa palautetta palkintojen tai rangaistusten muodossa. (Mohamed, 2025). Mohamedin (2025) mukaan

vahvistusoppiminen on lupaavaa erityisesti mukautuvien puolustusmenetelmien toteutuksessa. Vahvistusoppimista voidaan käyttää esimerkiksi tunkeutumisen estojärjestelmissä (intrusion prevention systems, IPS). Tunkeutumisen estojärjestelmässä malli oppii valitsemaan parhaan toimenpiteen kuhunkin hyökkäystyyppiin (Mohamed, 2025). Vahvistusoppimista voidaan myös hyödyntää myös automaattisessa penetraatiotestaamisessa, jossa tekoäly tutkii haavoittuvuuksia hallitussa ympäristössä ja oppii joko hyödyntämään tai torjumaan niitä. (Mohamed, 2025).

3.2.4 Syväoppiminen ja neuroverkot

Syväoppimisella (deep learning, DL) tarkoitetaan monikerroksisiin keinotekoisiin neuroverkkoihin (artificial neural networks) perustuvia menetelmiä, jotka pyrkivät mallintamaan monimutkaisia tietorakenteita (Halbouni ja muut, 2022; Mohamed, 2025). Perinteisiin koneoppimismenetelmiin verrattuna DL-malli vähentää manuaalista piirresuunnittelua, mutta vaatii silti esikäsittelyä (Halbouni ja muut, 2022). DL-mallit soveltuvat erinomaisesti pilvipalveluiden verkkoliikenneympäristöön, jossa data on runsasta ja jatkuvasti muuttuvaa.

Neuroverkkojen arkkitehtuurit vaihtelevat käyttötarkoituksen mukaan. Konvoluutioneuroverkot (convolutional neural networks, CNN) soveltuvat hyvin rakenteellisten piirteiden tunnistamiseen esimerkiksi verkkoliikenteen kuvioista (Mohamed, 2025). Toistuvat neuroverkot (recurrent neural network, RNN) kykenevät mallintamaan aikajonoihin perustuvia riippuvuuksia, kuten käyttäjien toiminnan ajallisia poikkeamia. Lisäksi generatiiviset vastakkaiset verkot (generative adversarial network, GAN), tarjoavat mahdollisuuksia sekä hyökkäysten simulointiin että puolustuksen vahvistamiseen tuottamalla realistista harjoitusdataa.

Mohamed (2025) korostaa, että syväoppimiseen pohjautuvilla tietoturvaratkaisuilla on merkittävä etu nollapäivähyökkäyksien havaitsemisessa, sillä neuroverkot voivat mukautua jatkuvasti muuttuviin uhkakuviin. Luonnollisesti syväoppimiseen liittyy myös haasteita, esimerkiksi mallien korkea laskennallinen kuormittavuus ja mahdollinen

ylisovittaminen (overfitting) voivat vaikeuttaa niiden käyttöä kriittisissä pilvipalveluympäristöissä (Halbouni ja muut, 2022).

3.3 Ennustavat mallit

Ennustavilla malleilla tarkoitetaan tässä tutkielmassa menetelmiä, jotka arvioivat tulevan poikkeaman tai riskitason todennäköisyyttä aikaleimatusta loki- ja verkkodatasta. Tavoite on proaktiivinen suojaus, eli varoitus annetaan ennen häiriön tai hyökkäyksen toteutumista. (Abdallah ja muut, 2024).

Tätä lähestymistä tukevat toistuvat neuroverkot (recurrent neural network, RNN), jotka on suunniteltu jono- ja aikasarjadataan. Ne oppivat aikariippuvuuksia muuttuvapituisten syötteiden yli (Dong ja Kotenko, 2025).

RNN-perheen tunnetuin alaosasto on pitkän lyhytaikaisen muistin verkko (long short-term memory, LSTM), jossa porttirakenne mahdollistaa pitkän aikavälin riippuvuuksien oppimisen (Dong ja Kotenko, 2025). Portitettu toistoyksikkö (gated recurrent unit, GRU) on kevyempi, mutta yksinkertaistaa portitusta ja säilyttää kyvyn mallintaa ajallista rakennetta. (Mohamed, 2025).

Autoenkooderi (autoencoder, AE) on pakkaava-purkava neuroverkko, jota käytetään valvomattomaan poikkeamien havaitsemiseen (Dong ja Kotenko, 2025). Malli opetetaan rekonstruoimaan normaalia käyttäytymistä ja rekonstruointivirhe toimii hälytyskriteerinä. Ennakoivassa käytössä hälytyksen voi muodostaa myös ennustevirhe, eli ero mallin ennusteen ja toteuman välillä (Abdallah ja muut, 2024).

3.4 UEBA ja Hybridimallit

UEBA (User and Entity Behavior Analytics) viittaa käyttäjien ja entiteettien toiminnan analysointiin. Sen avulla voidaan rakentaa profiili käyttäjän normaalista toimintamallista ja käyttää sitä epätavallisen toiminnan havaitsemiseen (Marchal ja muut, 2024). UEBA-

ratkaisut ovat luonteeltaan hybridejä, jotka yhdistävät tilastollisia peruslinjoja, ohjaamatonta oppimista ja tarvittaessa valvottuja tai DL-malleja riskipisteytyksen tuottamiseksi (Ahmetoglu & Das, 2022). Käytännössä yhdistämisstrategioita ovat kaksiportainen tunnistus, jossa ohjaamaton malli seuloo poikkeavat tapaukset valvotun luokittelun tarkennettavaksi, sääntö-/allekirjoituspohjaisen tunnistuksen ja ML-mallin yhdistäminen sekä graafi- ja sekvenssipohjaisten mallien yhdistelmät (Abdallah ja muut, 2024).

UEBA on erityisen hyödyllinen monivuokraisissa pilviympäristöissä, joissa kaapatut tilit ja API-väärinkäyttö eivät näytä allekirjoitushyökkäyksiltä. Hybridiratkaisut auttavat pitämään FP-tason hallinnassa ilman, että nollapäiväuhka lisääntyy (Marchal ja muut, 2024).

Tuore tutkimus osoittaa, että allekirjoituspohjaisen IDS:n rikastaminen ML- ja DL-tekniikoilla parantaa havaitsemistarkkuutta ja vähentää vääriä hälytyksiä, mikä vahvistaa järjestelmän hälytysvakautta dynaamisessa ympäristössä (Ahmed ja muut, 2025; Dong & Kotenko, 2025; Abdallah ja muut, 2024).

UEBA:n käytännön rajoitteena voi pitää sen tarvitsemaa pitkää oppimisjaksoa sekä tarvittavan historiallisen datan määrää (Marchal ja muut, 2024). Tämä voi heikentää suojaustasoa uusille käyttäjille ja tuotantokäytössä korostuu mallien säännöllinen päivitys.

4 Mallien arviointi ja soveltuvuus

Tässä luvussa arvioidaan AI- ja ML-menetelmiin perustuvia tietoturvaratkaisuja. Arviointikehikon tarkoitus on tehdä tuloksista vertailukelpoisia ja käytännössä tulkittavia. Valitut mittarit kuvaavat sekä havaintotarkkuutta että operatiivista käytettävyyttä viiveen, kuormituksen ja skaalautuvuuden osalta. Mittarivalinnat perustuvat alan vakiintuneisiin käytäntöihin (Mohamed, 2025; Dong ja Kotenko, 2025; Ahmetoglu ja Das, 2022; Nassif ja muut, 2021).

4.1 Arviointikehikko ja mittarit

Mohamedin (2025) mukaan luokittelumallin yleisen suorituskyvyn arvioinnissa Tarkkuus (Accuracy, Acc) on yksi perusmittareista. Tarkkuus määrittää oikein luokiteltujen entiteettien osuuden kaikkien arvioitujen entiteettien joukosta, eli korkea tulos viittaa mallin kykyyn luokitella haitallinen ja ei-haitallinen aktiivisuus. Tarkkuus voidaan laskea yksinkertaisella yhtälöllä:

$$\mathrm{Acc} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \quad (1)$$

missä

TP on oikein tunnistetut uhat (todet positiiviset, True Positives)

TN on oikein luokitellut ei-haitalliset toiminnot (todet negatiiviset, True Negatives)

FP on virheellisesti merkityt uhat

FN on todelliset uhat, joita malli ei onnistunut havaitsemaan (väärät negatiiviset, False Negatives)

Nämä neljä komponenttia perustuvat sekaannusmatriisiin, joka nähdään taulukossa 1.

Taulukko 1 Sekaannusmatriisi (Halbouni ja muut, 2022)

	Ennustettu: Positiivinen	Ennustettu: negatiivinen
Positiivinen	Tosi positiivinen (TP)	Väärä negatiivinen (FN)
Negatiivinen	Väärä positiivinen (FP)	Tosi negatiivinen (TN)

Mohamed (2025) toteaa myös, että todellisuudessa Acc ei yksin riitä kyberturvallisuuden monimuotoiseen ympäristöön, sillä tietoturvadata on epätasapainoista ja väärät hälytykset kuormittavat operatiivista toimintaa. Ahmetoglu ja Das (2022) korostavat usean mittarin tarpeellisuutta kyberturvallisuuteen tarkoitettujen mallien arvioinnissa. Arvioinnissa painotetaan sekä havaitsemista että väärien hälytysten hallintaa.

Täsmällisyys (Precision*, PPV)* kuvaa mallin kykyä välttää ei-haitallisten havaintojen virheellistä merkintää haitallisiksi; korkea PPV tarkoittaa vähäistä FP-määrä (Halbouni, 2022).

$$\mathrm{PPV} = \frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FP})} \tag{2}$$

Kattavuus tai todellisten positiivisten osuus (Recall, *TPR*) kuvaa mallin kykyä löytää kaikki uhat (Mohamed, 2025).

$$\mathrm{TPR} = \frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FN})} \tag{3}$$

F1-luku (F1) kuvaa harmonista keskiarvoa. Suurempi F1-luku on osoitus tasapainoisemmasta ja tehokkaammasta mallista. (Mohamed, 2025; Halbouni, 2022).

$$\mathrm{F1} = 2 \cdot \frac{\mathrm{PPV} \cdot \mathrm{TPR}}{\mathrm{PPV} + \mathrm{TPR}} \tag{4}$$

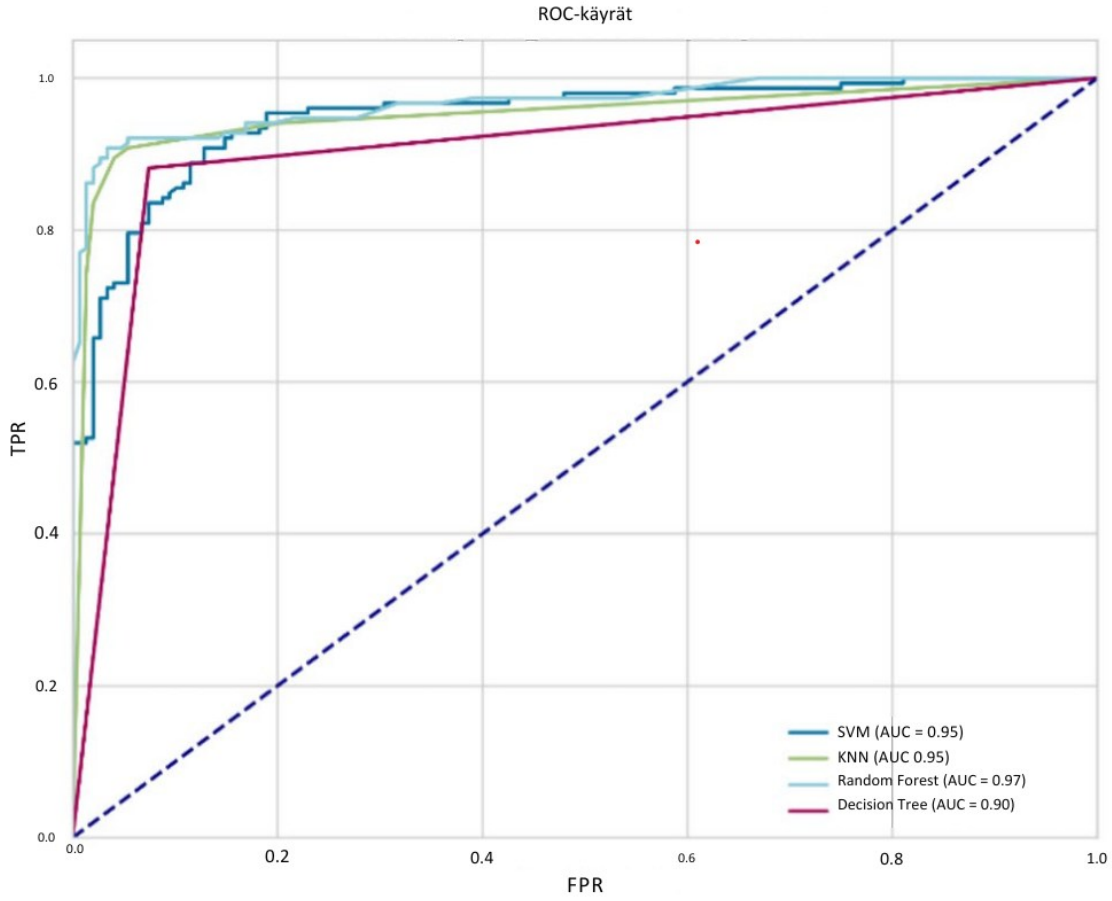
Väärien hälytysten osuus (False Positive Rate*, FPR*) mittaa, kuinka usein ei-haitallinen aktiivisuus luokitellaan haitalliseksi (Mohamed, 2025; Halbouni, 2022).

$$\mathrm{FPR} = \frac{\mathrm{FP}}{(\mathrm{FP} + \mathrm{TN})} \tag{5}$$

Väärien negatiivien osuus (False Negative Rate*, FNR*) mittaa, kuinka monta uhkaa malli epäonnistuu havaitsemaan. Alhainen FNR-lukema on välttämätön, jotta kyberturvajärjestelmä ei jätä kriittisiä hyökkäyksiä huomioimatta (Mohamed, 2025).

$$\mathrm{FNR} = \frac{\mathrm{FN}}{(\mathrm{FN} + \mathrm{TP})} \tag{6}$$

ROC-AUC (vastaanottajan toimintakäyrän alla oleva alue, Receiver Operating Characteristic - Area Under Curve) on laajalti käytetty mittari, joka kuvaa TPR:n ja FPR:n vaihtosuhdetta. AUC tiivistää tämän yhteen lukuun, jossa tulos lähellä lukua 1 kertoo erinomaisesta mallista ja vastaavasti tulos lähempänä lukua 0,5 kertoo epäluotettavasta mallista (Mohamed, 2025; Ahmetoglu ja Das, 2022). Kuvassa 3 on mukailtu eräitä ROC-käyriä (Ahmed ja muut, 2025). Kuvan mallit saavuttavat AUC-arvoja väliltä 0,90–0,97, mikä indikoi korkeaa erottelukykyä.



Kuva 3 Eräitä ROC-käyriä (Ahmed ja muut, 2025)

Inferenssiaika (inference time) kertoo, kuinka paljon aikaa mallilla menee analysoida ja luokitella mahdollisia turvauhkia. Myös prosessorin (Central Processing Unit, CPU), näytönohjaimen (GPU) ja keskusmuistin (RAM) laskenta- ja muistikuormaa tulee mitata. Mohamedin (2025) mukaan Tekoälypohjaisissa tietoturvaratkaisuissa on hyvä arvioida myös laskennallisen tehokkuuden perusteella, sillä mallien skaalautuvuus ja kustannukset ovat kriittisiä toiminnan tehokkuuden kannalta. Skaalautuvuutta mitataan myös tarkastelemalla, miten tekoälyohjelma sopeutuu kasvavaan kuormitukseen.

Seuraavaksi edellä mainittuja mittareita sovelletaan kolmeen yleisimpään käyttötilanneprofiiliin pilvi- ja tietokantaympäristössä, jotta mittarit kiinnittyvät konkreettisiin operatiivisiin vaatimuksiin.

4.2 Käyttötilanneprofiilit pilvessä ja tietokannoissa

Tässä alaluvussa rajataan tarkastelu kolmeen käyttötilanneprofiiliin, jotka toistuvat pilvi- ja tietokantaympäristöissä: (i) pilviverkon liikenteen valvonta, (ii) sovellus- ja PaaS-tason lokit sekä UEBA ja (iii) tietokantakyselyiden ja mallimuutosten poikkeavuudet. Profiilit peilautuvat CIA-luokitteluun: (1) saatavuus, (2) eheys ja (3) luottamuksellisuus.

Pilvipalvelujen moniasiakkuus ja skaalautuvuus siirtävät painopistettä volyymipoikkeamiin ja DDoS-hyökkäyksiin, joissa dataohjatut NIDS-ratkaisut (Network-based IDS) ovat luontevia. Abdallah ja muut (2024) osoittavat että DDoS on pilviturvallisuuden tutkituimpia teemoja; siksi tämä profiili asettaa reaaliaikaisuuden ja matalan FPR:n keskiöön.

UEBA hyödyntää valvontalokeja, käyttöoikeustapahtumia ja palvelukohtaisia mittareita sisäpiiriuhkien ja kaapatun tilin kaltaisten ilmiöiden havaitsemiseen. Ahmetoglu ja Das (2022) korostavat valvonnan siirtymistä pelkästä allekirjoituslogiikasta ennustaviin malleihin, mikä tukee sekä eheyden että luottamuksellisuuden suojaamista. Profiili suosii malleja, jotka sietävät harvinaisia luokkia ja muuttuvaa kontekstia, sekä edellyttää mittareiden tulkintaa operatiivisen hälytyskuorman kautta.

Tietokantatasolla hyökkäykset ilmentyvät mm. epätavallisina kyselyprofiileina, massapoimintoina tai injektioina. ML-menetelmiä on hyödynnetty sekä SQL-injektioiden havaitsemiseen että normaalin kyselykäyttäytymisen mallintamiseen siten, että poikkeavat rakenteet ja pääsykuviot paljastuvat. Tässä profiilissa eheyden ja luottamuksellisuuden vaatimukset korostavat matalaa FN-tasoa; pienikin sivuuttaminen voi johtaa tietovuotoon tai hiljaiseen manipulointiin (Ahmetoglu ja Das, 2022).

Profiileihin liittyy aineistoriippuvuus: yleisesti käytetyt IDS-aineiston, kuten KDD-perhe, UNSW-NB15, CIC-IDS eivät aina kata pilvipalveluiden nykyistä liikennettä tai organisaatiokohtaista kyselykäyttäytymistä, mikä vääristää mittareita ja siirrettävyyttä tuotantoon (Dong ja Kotenko, 2025).

4.3 Menetelmien soveltuvuus käyttötilanneprofiileihin

Mohamedin (2025) mukaan tunnettujen hyökkäysten havaitsemisessa ohjattu oppiminen toimii hyvin, kun käytössä on edustava ja ajantasainen verkkoa kuvaava opetusdata. Nollapäiväiset variantit, korkeadimensionaalinen virta ja vaihtuva kuormitus puoltavat ohjaamatonta anomaliantunnistusta sekä syväoppimista, kuten autoenkoodereita ja sekvenssimalleja. Katsaukset raportoivat, että DL parantaa havaitsemista ja voi alentaa FP määrää useissa vakiintuneissa dataseteissä, kuten CSE-CIC-IDS2018 ja CIC-IDS2017, mutta tulokset ovat datariippuvaisia. Pilvipainotteiset koonnit kokoavat samat havainnot IaaS- ja PaaS-ympäristöihin ja korostavat datasettien roolia. (Dong ja Kotenko, 2025).

Käyttäytymisen analytiikassa poikkeamat ovat usein harvinaisia ja kontekstiriippuvaisia, mikä suosii ohjaamatonta oppimista. Klusterointi ja autoenkooderit oppivat organisaatiokohtaisen normaalin ja hälyttävät siitä poikkeavat jaksot. Aikajonoihin perustuvat ilmiöt edellyttävät sekvenssimalleja, kuten RNN- ja LSTM-rakenteita. Kun laadukkaita luokiteltuja esimerkkejä on saatavilla, ohjattu luokittelu toimii tarkentavana toisena portaana. AI-vetoisen turvallisuusmallinnuksen yleiskatsaukset tukevat tätä kaksivaiheista virtausta ja kokoavat menetelmät loki-, käyttäjä- ja sovellusdatan päälle (Sarker ja muut, 2021).

Tietokantatasolla poikkeamat näkyvät epätavallisina kyselyprofiileina, massapoimintoina ja injektioina. Organisaatiokohtaisen normaalin hahmottamiseen soveltuu ohjaamaton oppiminen, kun taas tunnettujen hyökkäysmuotojen tunnistamiseen sopii ohjattu oppiminen. Laajat koonnit raportoivat ML- ja DL-menetelmiä injektioiden ja muun sovellustason poikkeavan käytöksen havaitsemiseen, mukaan lukien SQL-injektion valvonta neuroverkoilla ja hybridimalleilla (Ahmetoglu ja Das, 2022). Dong ja Kotenko (2025) viittaavat että hybridit yhdistävät havaitsemisherkkyyden ja maltillisen hälytyskuorman paremmin kuin yksittäiset mallit heterogeenisessa datassa. Syväoppiminen oppii esityksiä suoraan liikenne- ja lokivirroista ja voi parantaa tarkkuutta, mutta vaatii laskenta- ja

dataresursseja sekä huolellista siirrettävyyden arviointia. Sekä perinteiset ML-mallit että DL-mallit suoriutuvat vahvasti verkko-IDS-tehtävissä. (Dong ja Kotenko, 2025).

4.4 Aineistot ja siirrettävyys

ML-pohjaisten IDS-ratkaisujen arviointi nojaa usein julkisiin vertailuaineistoihin. Nassifin (2021) mukaan KDD ja NSL-KDD sisältävät ajallisesti vanhentuneita ja keskenään korreloivia piirteitä, mikä paisuttaa tuloksia eikä kuvaa pilvi- ja monivuokrausympäristöjen nykyistä liikennettä. Taulukko 2 kokoaa IDS-aineistojen peruspiirteet. Uudemmat CIC-IDS2017 ja CSE-CIC-IDS2018 laajentavat hyökkäyskirjoa, mutta eivät kata kaikkia kuormitus- ja palveluprofiileja, erityisesti anomaliatunnistuksessa normaaliluokan edustavuus on ratkaisevaa (Dong ja Kotenko, 2025). CIC-IDS2017:n hyökkäystyypit on esitetty taulukossa 3.

Taulukko 2 IDS-aineistojen yleiskuva (Halbouni ja muut, 2022)

Aineisto	Vuosi	Saatavuus	Piirteiden määrä	Liikenteen tyyppi
KDD Cup99	1998	Julkinen	41	Emuloitu
NSL-KDD	1998	Julkinen	41	Emuloitu
ISOT	2010	Julkinen	49	Emuloitu
ISCX 2012	2012	Julkinen	8	Emuloitu
UNSW-NB15	2015	Julkinen	42	Emuloitu
KYOTO	2015	Julkinen	24	Aito liikenne
CIC-IDS2017	2017	Julkinen	84	Emuloitu

Tuotantotulkinnassa aiemmin määritellyt mittarit tulee sitoa operatiiviseen kontekstiin. Epätasapainoisessa datassa ACC on heikko päätösperuste; tärkeämpiä ovat FNR, FPR

sekä latenssi ja läpivirtaus. Nämä määrittävät, muuttuuko oikea hälytys käytännössä myöhästyneeksi. (Mohamed, 2025).

Siirrettävyyden parantamiseksi Dong ja Kotenko (2025) suosittelevat datajoukon testaamista ristiin, aikajärjestyksen säilyttäviä jakoja kalibrointia ja konseptiajautumisen (concept drift) valvontaa. Näiden lisäksi vaiheittaista käyttöönottoa tulisi suosia. Nämä käytännöt kaventavat laboratorio-tuotanto-kuilua.

4.5 Käyttöönoton reunaehdot

Pelkkä mallitarkkuus ei riitä käyttökelpoisuuteen. Pilvi- ja tietokantaympäristöissä ratkaisevia ovat viive, läpivirtaus, kustannukset, integraatio, selitettävyys ja tietosuoja (Ahmetoglu & Das, 2022). Nämä tulee arvioida yhdessä havaitsemismittareiden kanssa, jotta tutkimustulokset siirtyvät tuotantokyvykkyydeksi.

Reaaliaikaisessa valvonnassa malli käsittelee suuria tapahtumamääriä pienellä viiveellä. DL-mallit kuormittavat laskentaa ja voivat kasvattaa viivettä, mikä lisää riskiä ohi menneistä havainnoista. Siksi arvioinnissa tulee raportoida sekä havaitsemismittarit että läpivirtaus ja vasteaika. Dong ja Kotenko (2025) nimeävät skaalautuvuuden, ajoitusvaatimukset ja väärien hälytysten hallinnan keskeisiksi käyttöönoton haasteiksi ja suosittelee tehokkuusvertailuja IDS/IPS-ratkaisuille.

Taulukko 3 CIC-IDS2017-aineiston hyökkäystyypit, (Halbouni ja muut, 2022)

Hyökkäysluokka	Tietueiden määrä	Kuvaus
Hyvälaatuinen	2 358 036	Normaali verkkoliikenne.
DoS -DDoS	41 835	Useat käyttäjät hyökkäävät samanaikaisesti yhtä palvelua vastaan.
DoS -Heartbleed	11	Luvaton pääsy syöttämällä haitallista dataa OpenSSL-muistiin.
DoS -Hulk	231 073	Hulk-työkalun tuottama obfuskettu liikenne DoS-hyökkäyksen toteuttamiseksi.
DoS -Slowloris	5 796	Slowloris-työkalulla toteutettu hidas yhteyshyökkäys.
PortScan	158 930	Tietojen keruu (palvelut, käyttöjärjestelmä) lähettämällä paketteja eri kohdeporteihin.
Web-hyökkäys -XSS	652	Haitallisen sisällön syöttäminen verkkosovellusten kautta tavallisille sivustoille.
Web-hyökkäys -murtoyritys (Brute Force)	1 507	Salasanojen arvaamiseen perustuvat murtoyritykset web-sovelluksissa.
Web-hyökkäys -SQL-injektio	21	Haitallisten SQL-lauseiden syöttäminen syötekenttiin suoritettavaksi.
Murtoyritys -FTP-Patator	7 938	Hyökkäykset FTP-kirjautumisen salasanan arvaamiseksi.
Murtoyritys -SSH-Patator	5 897	Hyökkäykset SSH-kirjautumisen salasanan arvaamiseksi.
Bot	1 966	Troijalainen kaappaa laitteita bottiverkoksi etäohjattavaksi.
Infiltration	36	Tunkeutumistekniikat ja -työkalut luvattoman pääsyn saamiseksi järjestelmädataan.

5 Eettiset ja lainsäädännölliset näkökulmat

Tässä luvussa tarkastellaan AI-pohjaisten poikkeamien havaitsemisratkaisujen eettisiä ja lainsäädännöllisiä reunaehtoja pilvi- ja tietokantaympäristöissä. Painopisteinä ovat (i) yksityisyys ja tietosuoja, (ii) vinoumat ja oikeudenmukaisuus, sekä (iii) selitettävyys ja vastuu.

5.1 Yksityisyys ja tietosuoja

Poikkeamien havaitseminen kyberturvallisuudessa nojaa yksityiskohtaiseen tapahtumadataan, joka herättää merkittäviä eettisiä huolenaiheita, erityisesti yksityisyyden, läpinäkyvyyden ja algoritmisen vinouman osalta (Mohamed, 2025). Valvonta- ja havaitsemisjärjestelmät vaativat laajaa pääsyä henkilötietoja sisältäviin loki- ja käyttäytymisaineistoihin, mikä korostaa sääntelyn noudattamisen tärkeyttä. IDS-toteutuksissa henkilötietojen tallentamista on vältettävä, jotta datan käyttö pysyy eettisenä. Tämä voidaan toteuttaa anonymisoimalla data ennen käyttöä (Ahmed ja muut, 2025).

5.2 Vinoumat ja oikeudenmukaisuus

Ahmed ja muut (2025) nostavat keskeisenä huolenaiheena koulutusdatasta perityt vinoumat. AI/ML-mallien tehokkuus perustuu käytössä olevaan koulutusdata, jos tämä aineisto sisältää vinoutunutta dataa, mallit perivät ja voivat vahvistaa koulutusdatan vinoumia. Kyberturvallisuudessa tämä voi johtaa joidenkin ryhmien suhteettomaan valvontaan tai harvinaisten uhkien alivalvontaan (Mohamed, 2025). Vinoumien pienentäminen edellyttää edustavia aineistoja, vinoumien mittaamista ja korjausmenetelmiä sekä ihmisen osallistamista päätöksiin, jotta kriittiset toimet vahvistetaan ennen toteutusta. Eettinen kustannus näkyy myös mittareissa: korkea FPR aiheuttaa hälytysväsymystä, kun taas FNR kasvattaa riskiä, joten kynnysarvot on kalibroitava riskiperusteisesti (Mohamed, 2025).

5.3 Selitettävyys ja vastuu

Syväoppimiseen perustuvat IDS-ratkaisut tarvitsevat selitettävyyttä (Explainable AI, XAI), jotta turvallisuustiimit ymmärtävät, miksi tapahtuma luokiteltiin uhaksi. (Mohamed, 2025.) Usein AI- ja etenkin ML-mallit toimivat mustan laatikon periaatteella, eli malli ei perustele ihmiselle miksi se on päätynyt kyseiseen ratkaisuun. Mohamedin (2025) mukaan XAI-malleille on kasvava kysyntä. Käytännössä tämä tarkoittaa päätöksen perustelevia malleja, sekä dokumentoituja perusteluja ja jäljitettävyyttä. Näiden ongelmien ratkaisemiseksi on kehitetty esimerkiksi LIME (Local interpretable Model-agnostic Explanations) ja SHAP (SHapley Additive, exPlanations) (Ahmed ja muut, 2025). LIME luo yksittäisille ennusteille perusteluita approksimoimalla mallin tekemiä päätöksiä mustan laatikon ympäristössä. SHARP perustelee ennusteita peliteoriaa (game theory) hyödyntäen.

6 Johtopäätökset

Tässä tutkielmassa vertailtiin AI/ML-pohjaisia poikkeamien havaitsemismalleja pilvi- ja tietokantaympäristössä. Tutkimuksen pääasiallisena tutkimuskysymyksenä oli ”Kuinka tekoälypohjaiset poikkeamien havaitsemismallit voivat edistää tietoturvaa pilvi- ja tietokantaympäristöissä?” Lisäksi avustavina tutkimuskysymyksinä oli ”Mitä eroja tunnistettujen mallien välillä on tehokkuuden, tarkkuuden ja soveltuvuuden näkökulmasta pilvi- ja tietokantojen tietoturvaympäristössä?” ja ”Mitä käytännön haasteita ja vaatimuksia liittyy tekoälypohjaisten poikkeamien havaitsemismallien käyttöönottoon pilvi- ja tietokantaympäristössä?”

AI/ML-pohjaiset poikkeamien havaitsemismenetelmät tarjoavat käytännön hyötyä erityisesti tilanteissa, joissa hyökkäys ei vastaa tunnettuja allekirjoituksia. Ne parantavat havaitsemisherkkyyttä ja lyhentävät reagointiaikaa pilviympäristöissä, joissa kuormat ja liikennemallit vaihtelevat nopeasti. Paras tasapaino herkkyyden ja väärien hälytysten välillä saavutetaan hybridiratkaisuilla: ohjaamaton anomaliantunnistus karsii poikkeamat ja valvottu tai DL-malli tarkentaa päätöksen; useissa pilvipainotteisten katsauksissa hybridien todettiin parantavan sekä tarkkuutta että FP-osuutta (Abdallah, 2024).

Tietokantaympäristöjen anomaliantunnistusta koskeva vertaisarvioitu näyttö on selvästi niukempaa kuin verkko-IDS-tutkimuksessa. Avoimien, realististen kyselylokienvähyys ja niiden yksityisyysrajoitteet heikentävät tulosten vertailtavuutta ja siirrettävyyttä. Siksi raportointi tulisi sitoa ympäristö- ja datakontekstiin, käyttää ajallisesti realistisia jakoja ja ristiintestausta sekä täydentää arviointia syntetisoiduilla tai anonymisoiduilla lokeilla.

Käyttöönoton kannalta pelkkä tarkkuus ei riitä, vaan arviointi on sidottava operatiivisiin mittareihin: FP, FN, läpivirtaus, latenssi, inferenssiaika ja resurssikuorma ratkaisevat, toteutuuko oikea havainto ajallaan ja kohtuullisella kustannuksella.

Mittariston on katettava FPR, FNR, AUC, F1 sekä muutokset ajan yli; tämä on vakiintunut lähtökohta IDS-tutkimuksessa. Tulosten siirrettävyyttä rajoittavat datalähteet: KDD/NSL-

KDD ovat ajallisesti ja rakenteellisesti vanhentuneita, kun taas CICIDS2017 ja CSE-CIC-IDS2018 laajentavat hyökkäyskirjoa, mutta eivät vielä kuvaa kaikkia pilvikuormia tai hyvänlaatuisen käyttäytymisen vaihtelua. Tästä syystä ristiintestaus, aikajärjestystä säilyttävät jaot ja konseptiajautumisen valvonta ovat suositeltavia. Syväoppiminen voi nostaa peittoa ja tarkkuutta, mutta kasvattaa laskentakuormaa ja on siksi punnittava viive- ja skaalautuvuusvaatimuksia vasten; pilvessä korostuu lisäksi tarve skaalautuville ja yksityisyystietoisille ratkaisuille.

Jatkotutkimus tulisi kohdentaa mallien siirrettävyyteen ja ajallisten muutosten hallintaan pilvi- ja tietokantaympäristöissä. Erityistä huomiota tulee kiinnittää kyselyprofiilien ja käyttöoikeuskuvioiden muuntumiseen skaalautuvissa ja monivuokratuissa järjestelmissä. Tarvitaan evaluointipenkkejä, jotka huomioivat luokkien epätasapainon ja mahdollistavat vertailun sekä vakiintuneisiin IDS-aineistoihin, että organisaatiokohtaisiin tietokantalokeihin; arvioinnin tulee kattaa myös salattu liikenne ja tiukat tietosuojavaatimukset.

Lähteet

Abdallah, A. M., Alkaabi, A. S. R. O., Alameri, G. B. N. D., Rafique, S. H., Musa, N. S., & Murugan, T. (2024*). Cloud Network Anomaly Detection Using Machine and Deep Learning Techniques—Recent Research Advancements*. *IEEE Access*, *12*, 56749-56773. https://doi.org/10.1109/ACCESS.2024.3390844

Ahmed, U., Nazir, M., Sarwar, A., Ali, T., Aggoune, E.-H. M., Shahzad, T., & Khan, M. A. (2025). *Signature-based intrusion detection using machine learning and deep learning approaches empowered with fuzzy clustering*. *Scientific Reports*, *15*(1), 1726. https://doi.org/10.1038/s41598-025-85866-7

Ahmetoglu, H., & Das, R. (2022). *A comprehensive review on detection of cyber-attacks: Data sets, methods, challenges, and future research directions*. *Internet of Things*, *20*, 100615. https://doi.org/10.1016/j.iot.2022.100615

Aksela, M., Marchal, S., Patel, A., & Rosenstedt, L. (2022). *Tekoälyn mahdollistamat kyberhyökkäykset*. Liikenne- ja viestintävirasto Traficom. Noudettu 12. toukokuuta 2025 osoitteesta https://www.kyberturvallisuuskeskus.fi/fi/julkaisut/tekoalyn-mahdollistamat-kyberhyokkaykset

Dong, H., & Kotenko, I. (2025). *Cybersecurity in the AI era: Analyzing the impact of machine learning on intrusion detection*. *Knowledge and Information Systems*, *67*(5), 3915-3966. https://doi.org/10.1007/s10115-025-02366-w

Halbouni, A., Gunawan, T. S., Habaebi, M. H., Halbouni, M., Kartiwi, M., & Ahmad, R. (2022). *Machine Learning and Deep Learning Approaches for CyberSecurity: A Review. IEEE Access*, *10*, 19572-19585. https://doi.org/10.1109/ACCESS.2022.3151248

Marchal, S., & Nawrotek, B. (2024). *Tekoälypohjaiset kyberturvallisuusratkaisut*. Liikenne- ja viestintävirasto Traficom & WithSecure. Tekoälypohjaiset kyberturvallisuusratkaisut. Noudettu 12. toukokuuta 2025, osoitteesta

https://www.kyberturvallisuuskeskus.fi/sites/default/files/media/file/Teko%C3%A4lypohjaiset%20kyberturvallisuusratkaisut_FI.pdf

Mohamed, N. (2025). *Artificial intelligence and machine learning in cybersecurity: A deep dive into state-of-the-art techniques and future paradigms*. *Knowledge and Information Systems*, *67*(8), 6969-7055. https://doi.org/10.1007/s10115-025-02429-y

Nassif, A. B., Talib, M. A., Nasir, Q., Albadani, H., & Dakalbab, F. M. (2021). Machine Learning for Cloud Security: A Systematic Review. *IEEE Access*, *9*, 20717-20735. https://doi.org/10.1109/ACCESS.2021.3054129

Nasteski, V. (2017). An overview of the supervised machine learning methods. *HORIZONS.B*, *4*, 51-62. https://doi.org/10.20544/horizons.b.04.1.17.p05

Sarker, I. H., Furhad, M. H., & Nowrozy, R. (2021). AI-Driven Cybersecurity: An Overview, Security Intelligence Modeling and Research Directions. *SN Computer Science*, *2*(3), 173. https://doi.org/10.1007/s42979-021-00557-0

Zhengbing, H., Gnatyuk, S., Koval, O., Gnatyuk, V., & Bondarovets, S. (2017). *Anomaly Detection System in Secure Cloud Computing Environment*. *International Journal of Computer Network and Information Security*, *9*(4), 10-21.

https://doi.org/10.5815/ijcnis.2017.04.02

Solin, A. (2022). *Tekoälyratkaisut tänään ja tulevaisuudessa*. s. 69–73. Noudettu 24. heinäkuuta 2025, osoitteesta https://www.eduskunta.fi/FI/valiokunnat/tulevaisuusvaliokunta/julkaisut/Sivut/tekoalyratkaisut-tanaan-ja-tulevaisuudessa.aspx

IBM. (2021, 22. syyskuuta). *What Is Machine Learning?* Noudettu 24. heinäkuuta 2025, osoitteesta https://www.ibm.com/think/topics/machine-learning