

Koneoppiminen Azure Machine Learning -palvelussa

LAB-ammattikorkeakoulu
Tradenomi (AMK), Liiketalous
2021
Miia Romanainen

Tiivistelmä

Tekijä(t) Romanainen, Miia	Julkaisun laji Opinnäytetyö, AMK Sivumäärä 39	Valmistumisaika 2021
Työn nimi Koneoppiminen Azure Machine Learning -palvelussa		
Tutkinto Tradenomi (AMK)		
Toimeksiantajan nimi, titteli ja organisaatio LAB-ammattikorkeakoulu Oy		
Tiivistelmä <p>Opinnäytetyön tarkoituksena oli tutkia koneoppimista Azure Machine Learning -palvelussa ja selvittää palvelun hyödyntämistä tietojenkäsittelyn opetuksessa. Suuressa osassa viimeisimpiä tutkimuksia on vertailtu eri koneoppimista tarjoavia pilvipalveluja. Koska koneoppiminen on aiheena erittäin laaja, syvällisempi selvittäminen sen toteuttamisesta yksittäisessä palvelussa oli tarpeen.</p> <p>Työ toteutettiin toiminnallisena opinnäytetyönä. Teoriaosuudessa tarkasteltiin koneoppimista käsitteenä ja palvelua teknologian näkökulmasta. Toiminnalliset kokeilut käsittivät koneoppimismallin kehittämisen Azure Machine Learning Studioissa aina palvelun käyttöönotosta mallin arviointiin.</p> <p>Toiminnallisen osuuden tuloksena syntyi vaiheittainen ohjeistus Azure Machine Learning -ympäristön käyttöönottamiseksi, koneoppimismallin kehittämiseksi ja datan analysoimiseksi. Työn lopputuloksena saadaan kokemuseräistä tietoa ja syvällisempää ymmärrystä palvelusta ja sen käytettävyydestä opetuksessa.</p>		
Asiasanat koneoppiminen, Azure Machine Learning, pilvipalvelut		

Abstract

Author(s) Romanainen, Miia	Type of Publication Thesis, UAS	Published 2021
	Number of Pages 39	
Title of Publication Machine Learning in Azure Machine Learning Service		
Name of Degree Bachelor of Business Administration (UAS)		
Name, title and organization of the client LAB University of Applied Sciences		
Abstract <p>The purpose of this thesis was to investigate machine learning in the Azure Machine Learning service and its use in the teaching of information technology. Many recent studies have compared different cloud services offering machine learning. As machine learning is a very broad topic, a more in-depth study of its implementation in a single service was needed.</p> <p>A functional method was used in the thesis. In the theoretical part, machine learning as a concept and service from a technology perspective was studied. The functional experiments included the development of a machine learning model in Azure Machine Learning Studio, implementing the service and evaluating the model.</p> <p>As a result of the functional part, a step-by-step guidance for implementing the Azure Machine Learning environment, developing a machine learning model, and analysing data was created. The thesis offers experiential information and a deeper understanding of the service and its usability in teaching.</p>		
Keywords machine learning, Azure Machine Learning, cloud services		

Sisällys

1	Johdanto.....	1
1.1	Taustatietoa.....	1
1.2	Opinnäytetyön tarkoitus, tavoite ja rajaukset.....	2
1.3	Tutkimusmenetelmä ja rakenne.....	3
2	Koneoppiminen.....	4
2.1	Koneoppiminen käsitteenä.....	4
2.2	Koneoppimisen algoritmit.....	5
2.2.1	Ohjattu oppiminen.....	7
2.2.2	Ohjaamaton oppiminen.....	8
2.2.3	Vahvistusoppiminen.....	9
2.3	Koneoppimismallin arviointi.....	11
2.4	Koneoppiminen pilvipalveluna.....	11
3	Koneoppimisen sovelluskohteet.....	13
3.1	Yritysten sovelluskohteet.....	13
3.2	Esineiden internet.....	13
3.3	IoT ja Azure Machine Learning.....	15
4	Azure Machine Learning.....	17
4.1	Ominaisuudet.....	17
4.2	Azure Machine Learning Studio.....	18
4.3	Koneoppimisprojektin työnkulku.....	18
5	Azure Machine Learning käytännössä.....	21
5.1	Palvelun käyttöönotto.....	21
5.2	Mallin kehittäminen suunnittelutyökalulla.....	24
5.3	Datan tuonti.....	26
5.4	Datan valmistelu.....	27
5.5	Mallin kouluttaminen.....	29
5.6	Mallin arviointi.....	30
6	Yhteenveto ja pohdinta.....	34
	Lähteet.....	36

Liitteet

Liite 1. Machine Learning Algorithm Cheat Sheet

TERMIT

AI	<i>Artificial Intelligence</i> , tekoäly
API	<i>Application Programming Interface</i> , ohjelmointirajapinta
AutoML	<i>Automated Machine Learning</i> , automaattinen koneoppiminen
Azure Machine Learning	Microsoft Azuren pilvipalvelukomponentti koneoppimiseen, jonka käyttöliittymänä toimii Azure Machine Learning Studio.
CLI	<i>Command Line Interface</i> , komentoliittymä
IaaS	<i>Infrastructure as a Service</i> , infrastruktuuri palveluna
IIoT	<i>Industrial Internet of Things</i> , teollinen esineiden internet
IoT	<i>Internet of Things</i> , esineiden internet
Klusterointialgoritmi	<i>Clustering Algorithm</i>
Kokonaisratkaisu	<i>End-to-End Solution / Engineering</i>
Luokitin	<i>Classifier</i>
Luokitusalgoritmi	<i>Classification Algorithm</i>
ML	<i>Machine Learning</i> , koneoppiminen
MLaaS	<i>Machine learning as a service</i> , koneoppiminen palveluna
MLOps	<i>Machine Learning Operations</i> , koneoppimistoiminnot
Ohjaamaton oppiminen	<i>Unsupervised Learning</i>
Ohjattu oppiminen	<i>Supervised Learning</i>
Opetusdata	<i>Labeled Data</i>
PaaS	<i>Platform as a Service</i> , alusta palveluna
Regressioalgoritmi	<i>Regression Algorithm</i>
SaaS	<i>Software as a Service</i> , ohjelmiston toimittaminen palveluna
Vahvistusoppiminen	<i>Reinforcement Learning</i>
Yhdistetty todellisuus	<i>Mixed Reality</i>
Älykäs pilvi ja -ulkoreuna	<i>Intelligent Cloud and Intelligent Edge</i>

1 Johdanto

1.1 Taustatietoa

Koneoppiminen jatkaa yhä tärkeänä trendinä tietojenkäsittelyssä. Pilvipalvelut taas ovat keskeisessä asemassa koneoppimisessa, koska koneoppimismallien kehittäminen vaatii usein paljon laskentatehoa. Idea opinnäytetyön aiheesta tuli LAB-ammattikorkeakoulusta, tarpeista saada tietoa koneoppimisesta Azuren pilvipalvelussa ja sen hyödyntämistä opetuksessa. Aiheen valintaan vaikutti myös oma kiinnostukseni koneoppimiseen ja tutustuminen Azuren oppimisen työkalujen käyttöön opintojen aikana. Vaikka koneoppimisen historia ulottuu jo 1950-luvulle, aiheen ajankohtaisuudesta kertoo, että siitä löytyy paljon tuoreita tutkimuksia ja artikkeleita. Suuressa osassa viimeisimpiä tutkimuksia on vertailtu eri koneoppimista tarjoavia pilvipalveluja. Koska koneoppiminen on aiheena erittäin laaja, selvittäminen sen toteuttamisesta yksittäisessä palvelussa on myös tarpeen.

Koneoppiminen on kiinnostava liiketalouden näkökulmasta, sillä sitä voidaan hyödyntää monin eri tavoin liiketoiminnan tarpeisiin. Sovelluskohteita ovat esimerkiksi hakukoneet, puheentunnistus, konenäkö sekä ennustavat työkalut raakadatasta ja analyyseistä. Tekoäly ja koneoppiminen muuttavat työntekotapoja ja menetelmät ovat kehittyneet ja kehittyvät yhä nopeasti. Lähitulevaisuudessa koneoppimisen sovelluksista tulee monelle arkea työssä. Liiketoiminnassa muutoksia on odotettavissa ainakin kolmella eri osa-alueella: tehtävissä, liiketoimintaprosesseissa sekä liiketoimintamalleissa. Koneoppiminen täydentää ihmistoimintoja ja rutiinomaiset tehtävät pyritään usein automatisoimaan koneoppimisen ja muun tekoälyn avulla. (Brynjolfsson & McAfee 2017, 4–11.)

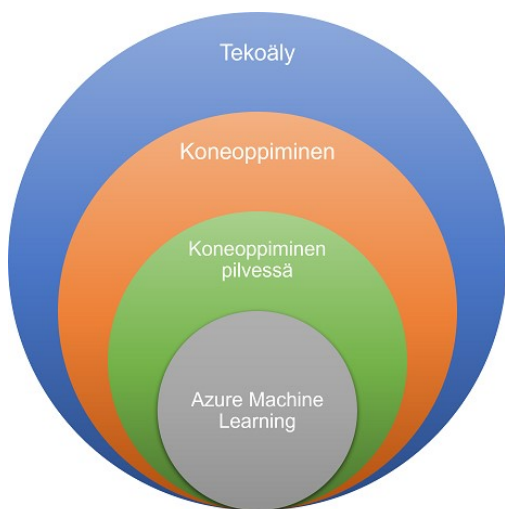
Azure Machine Learning on Microsoftin tarjoama pilvipalvelukomponentti koneoppimiseen, jonka käyttöliittymänä toimii Azure Machine Learning Studio. Palvelua voidaan käyttää koneoppimismallien opettamiseen, kehittämiseen, automatisointiin, hallintaan ja seurantaan. Se yhdistää yksinkertaisen koodillisen ja täysin koodittoman tavan koneoppimismallien käytölle. (Microsoft 2021b.) Palvelun kiinnostavuutta lisää se, että Gartnerin 2021 Magic Quadrant for Data Science and Machine Learning Platforms -julkaisun mukaan Microsoft on visionäärisin toimittaja Googlen ja AWS:n ohella. Microsoftilla on kyseisen luokan korkein suoritustaso ja se tarjoaa monipuolisesti mahdollisuuksia eri kohderyhmille kuten datatieteilijöille, kehittäjille ja data-analytiikoille. Sillä on myös vahvat MLOps-toiminnot, tietoturva sekä laskentakiintiöiden ja kustannusten hallinta. (King 2021.) Vastaavassa Gartnerin Cloud AI Developer Services -julkaisussa Microsoft on nimetty johtavaksi palveluiden toimittajaksi. Arvioitavia olivat palvelut erityisesti liittyen kieleen, konenäköön ja automatisoituun koneoppimiseen, joihin lukeutuu muun muassa Azure Machine Learning. (Baker ym. 2021.)

1.2 Opinnäytetyön tarkoitus, tavoite ja rajaukset

Opinnäytetyössä tarkastellaan koneoppimista Azure Machine Learning -palvelussa ja selvitetään palvelun hyödyntämistä tietojenkäsittelyn opetuksessa. Työssä pyritään vastaamaan kysymyksiin, mitä Azure Machine Learning on, mitä palvelun käyttö edellyttää ja miten palvelua voidaan hyödyntää opetuksessa. Tutkimuskysymyksiin pyritään löytämään vastaukset tutkimalla koneoppimista ja Azure Machine Learning -palvelua sekä tutustumalla niitä koskevaan aiempaan tutkimukseen ja kirjallisuuteen. Työssä käytetään hyödyksi myös Microsoftin omia aineistoja ja oppaita.

Opinnäytetyön teoriaosuudessa tavoitteena on tutkia koneoppimista käsitteenä, Azure Machine Learning -palvelua ja teknologian keskeisimpiä ominaisuuksia. Toiminnallisessa osuudessa tavoitteena on selvittää vaiheet ympäristön käyttöönotosta valmiin opetetun koneoppimismallin aikaansaamiseksi. Lisäksi selvitetään, miten data tuodaan palveluun sekä miten sitä voidaan hyödyntää, analysoida ja visualisoida. Tarkoituksena on saada kokemukseräistä tietoa palvelun käyttöönotosta ja käytettävyydestä opetuksessa.

Työ rajataan koneoppimisen ja pilvipalveluiden käsittelyyn teoreettisen viitekehyksen luomiseksi. Työn ylimmällä tasolla tarkastellaan koneoppimista tekoälyn osa-alueena. Koneoppimista käsitellään ainoastaan teknologian näkökulmasta ja työssä ei oteta kantaa koneoppimisen eettisiin tai pilvipalveluiden tietoturvaan liittyviin kysymyksiin. Työhön valitaan tarkastelun kohteeksi vain Azure Machine Learning -palvelu ja muiden pilvipalveluiden syvällisempää tarkastelua tai vertailua ei tehdä. Lisäksi työssä ei käsitellä opetusta muutoin kuin siinä tutkittavan palvelun hyödynnettävyyden kannalta. Työn rajaukset on esitetty kuvassa 1.



Kuva 1. Työn rajaukset

1.3 Tutkimusmenetelmä ja rakenne

Työ toteutetaan toiminnallisena opinnäytetyönä, jonka tavoitteena on käytännön toiminnan ohjeistus, opastus, toiminnan järjestäminen tai järjeistäminen. Lopputuloksena on itse raportti ja jokin produkti, joka voi esimerkiksi olla opas, ohjeistus tai suunnitelma. (Airaksinen 2009.) Vaikka toiminnallisessa opinnäytetyössä on tutkiva ja kehittävä ote, tutkimus on lähinnä selvityksen tekemistä tiedonhankinnan apuvälineenä (Vilkkä & Airaksinen 2003, 57).

Työn toiminnallisen osuuden tavoitteena on selvittää vaiheet Azure Machine Learning -ympäristön käyttöönotosta valmiin opetetun koneoppimismallin aikaansaamiseksi, datan analysoimiseksi ja visualisoimiseksi. Aineisto käsittää toimeksiantajan kanssa mallia varten valitun datan. Työn lopputuloksena syntyy selvitys Azure Machine Learning -palvelusta ja ohjeistus sen käytöstä koneoppimisen suorittamiseen.

Toimeksiantajana on LAB-ammattikorkeakoulu, joka toimii Lappeenrannassa, Lahdessa ja verkossa. Opintotarjontaan kuuluu sosiaali- ja terveysala, tekniikka, liiketalous, hotelli-, ravintola- ja matkailuala sekä muotoilu, taide ja visuaalinen viestintä. LAB-innovaatiokorkeakoulu aloitti toimintansa, kun Lahden ja Saimaan ammattikorkeakoulut yhdistyivät 1.1.2020. LAB nimenä ilmentää työelämän innovaatiokorkeakoulun visiota ja toimintakulttuuria. Se kuvaa myös aktiivista tekemistä, jossa erilaiset ihmiset ja osaajat luovat uutta yhdessä yritysten kanssa. (LAB-ammattikorkeakoulu 2021.)

Työ rakentuu kuudesta luvusta. Johdannossa (luku 1) kerrotaan työn taustasta, tavoitteista, rajauksista ja toteutusmenetelmästä. Teoria ja tietoperusta (luvut 2–5) käsittää koneoppimisen, koneoppimisen sovelluskohteet ja Azure Machine Learning -palvelun. Azure Machine Learning käytännössä (luku 5) käsittelee koneoppimismallin kehittämistä Azure Machine Learning Studioissa aina palvelun käyttöönotosta mallin arviointiin. Lopuksi yhteenveto ja pohdinta (luku 6) tarkastelee työn tuloksia, arvioi asetettujen tavoitteiden saavuttamista ja jatkokehittämismahdollisuuksia.

2 Koneoppiminen

2.1 Koneoppiminen käsitteenä

Koneoppimiselle ja tekoälylle löytyy kirjallisuudesta useita määritelmiä, ja ne eivät ole yksiselitteisiä. Koneoppiminen on tekoälyn osa-alue. Merilehdon (2018, 18) mukaan tekoälyllä tarkoitetaan useimmiten koneen suorittamaa toimintaa, joka ihmisen tekemänä olisi älykästä. Eräs tapa määritellä tekoäly on luetella sille tyypillisiä ominaisuuksia, joita ovat autonomisuus ja adaptiivisuus. Autonomisuudella tarkoitetaan kykyä suorittaa tehtäviä monimutkaisessa ympäristössä ilman jatkuvaa käyttäjän ohjausta. Adaptiivisuus taas on kyky parantaa suorituskkyä oppimalla kokemuksesta. (Helsingin yliopisto & Reaktor Innovations 2021.)

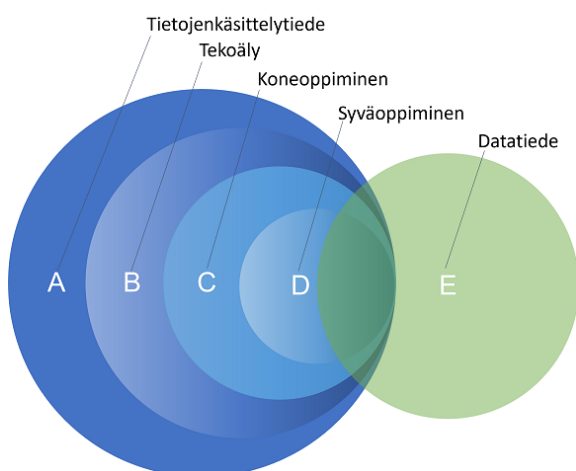
Tekoälyn avulla suoritettuja toimintoja ovat muun muassa päättely, oppiminen, ennakointi, päätöksenteko, näkö ja kuulo. Heikko tekoäly kykenee ratkaisemaan yhtä sille opetettua tehtävää kuten tunnistamaan konenäön avulla tiettyjä asioita kuvista. Vahva tekoäly sen sijaan esimerkiksi ratkoo laajasti erilaisia ongelmia, ajaa autoa, ymmärtää kieliä ja kokkaa. Nykyiset tekoälysovellukset ovat käytännössä heikkoa tekoälyä. (Merilehto 2018, 18–19.)

Suurin osa tekoälyn sovelluksista on koneoppimista. Koneoppimisella tarkoitetaan useissa lähteissä algoritmien kykyä oppia yleistettäviä malleja niin, että niitä on erikseen ohjelmoitu. Näitä malleja voidaan käyttää ennustamiseen, tietojen luokitteluun ja kuvailemiseen. (Merilehto 2018, 19; Microsoft 2021c.) Koneoppiminen on järjestelmien kykyä parantaa suorituskkyään tiettyssä tehtävässä sitä mukaa kuin lisää kokemusta tai dataa kertyy (Helsingin yliopisto & Reaktor Innovations 2021). Koneet oppivat siten ympäristöstään ja pystyvät tekemään itsenäisesti parempia ratkaisuja kuin pelkällä etukäteen tehdyllä ohjelmoinnilla (Collin & Saarelainen 2016, 210). Tekoälyn edelläkävijän Arthur Lee Samuelin 1950-luvulla antaman klassisen määritelmän mukaan koneoppiminen on joukko menetelmiä, jotka antavat tietokoneille *kyvyn oppia ilman, että niitä on erityisesti ohjelmoitu* (Samuel 2000). Koneoppimisalgoritmi siis löytää ja muodostaa periaatteita suuresta joukosta dataa. Siten algoritmi voi perustella ominaisuuksia datasta. Näitä matemaattisesti muodostettuja periaatteita kutsutaan malliksi. (Kaspersky Lab 2020, 3.)

Koneoppiminen on tiede koneiden harjoittamisesta ihmisten tapaan analysoimaan ja oppimaan tiedosta. Koneoppiminen on tekoälykäsitteen alle kuuluva prosessi, jossa matemaattisia malleja käytetään auttamaan tietokonetta oppimaan ilman suoraa ohjausta. (Microsoft 2021a.) Sen voidaan myös ajatella olevan pohjimmiltaan tiedon eristämistä datasta, koska koneoppimisen juuret ovat tilastotieteessä. Edelleen koneoppimisen ytimessä olevat tekniikat, kuten lineaariregressio ja bayesiläiset menetelmät, joista monet ovat jo yli kaksi

vuosisataa vanhoja. Erona perinteiseen tekoälyyn on se, ettei koneoppimisen avulla pyritä ihmisen älykkääseen käytökseen. Sen sijaan etsitään isosta data-aineistosta ne säännönmukaisuudet, joita ihminen ei löytäisi. Useimmissa koneoppimisen ongelmissa onkin vain yksi oikea vastaus. (Helsingin yliopisto & Reaktor Innovations 2021; itewiki 2021.) Erona tilastotieteeseen on taas se, että koneoppiminen perustuu ohjelmitaviin algoritmeihin ja vaikeisiin päättelyongelmiin. Keskeistä on kirjoittaa tutkittavaa dataa vastaan ajettavia algoritmeja, joiden avulla algoritmit oppivat datasta. (Collin & Saarelainen 2016, 210.)

Koneoppimisen asemoimiseksi voidaan hyödyntää Euler-kaaviota (Kuva 2). Kaikki tietojenkäsittelytieteestä syväoppimiseen (A–D) ovat osa datatiedettä (E). Koska datatiede tarvitsee tietojenkäsittelytieteen ja tekoälyn lisäksi myös muita tieteitä ja monien sovellusalueiden osaamista, sitä ei yleensä lasketa tietojenkäsittelytieteen osa-alueeksi (Helsingin yliopisto & Reaktor Innovations 2021).



Kuva 2. Koneoppimisen asemointi (mukailtu Helsingin yliopisto & Reaktor Innovations 2021)

2.2 Koneoppimisen algoritmit

Jotta koneoppiminen on mahdollista, luodaan sille algoritmit eli ohjelmoidut käskysarjat. Ne seuraavat mekaanisesti ohjelmoijan sille antamia käskyjä. Algoritmit voidaan toteuttaa monilla eri ohjelmointikielillä, sillä niiden kuvaamat ideat ja toimintatavat eivät riipu toteutettavasta kielestä. (Kokkarainen & Ala-Mutka 2002, 15.) Koneoppimisalgoritmi on koodia, joka auttaa tutkimaan, analysoimaan ja löytämään merkityksen monimutkaisissa tietojoukoissa. Jokainen algoritmi on rajattu joukko selkeitä vaiheittaisia ohjeita, joita kone seuraa tietyn tavoitteen saavuttamiseksi. Tavoitteena on luoda tai löytää malleja, joiden avulla voidaan

tehdä ennusteita tai luokitella tietoja. Koneoppiminen siis käyttää algoritmeja yhtäläisyyksien tunnistamiseen datassa, minkä pohjalta voidaan laatia ennustemalleja. (Microsoft 2021c.) Koneoppisen käyttämät algoritmit oppivat käytössä olevasta datasta askel askeleelta. Koneoppimisen malli kehittyy siten ja pystyy kuvailemaan käytettävissä olevaa dataa entistä tarkemmin ja ennustamaan lopputuloksia. Malli pystyy luomaan sitä tarkemman kuvan, mitä enemmän sillä on dataa käytettävissään. Mallin harjoittamisen jälkeen sille annetaan syöte, josta malli antaa tulokset. (Merilehto 2018, 28.)

Koneoppimisalgoritmit käyttävät parametreja, jotka perustuvat harjoitusdataan. Kun harjoitusdata laajentuu edustamaan maailmaa realistisemmin, algoritmi laskee tarkemmat tulokset. Lisääntyneen datan ja toistojen, kokemuksen ansiosta koneoppimisen tulokset ovat tarkempia, kuten ihmisetkin pääsevät parempiin tuloksiin harjoittelemalla. Koneoppimisen sopeutumiskyky tekee siitä loistavan valinnan tilanteissa, joissa data muuttuu jatkuvasti, pyynnön tai tehtävän luonne vaihtuu toistuvasti tai koodaamalla ratkaisu olisi käytännössä mahdotonta toteuttaa. (Microsoft 2021c.)

Eri algoritmit analysoivat tietoja eri tavoin. Usein ne ryhmitellään käytettyjen koneoppimistekniikoiden mukaan ohjatun oppimisen-, valvomattoman oppimisen- ja vahvistavaan oppimisen menetelmiin. (Microsoft 2021c.) Ohjatussa oppimisessa koneelle annetaan oikea vastaus opetusdatassa, kun ohjaamattomassa oppimisessä kone pääättelee asioita datassa olevien säännönmukaisuuksien ja suhteiden pohjalta. Vahvistusoppimisessä koneelle taas annetaan palautetta sen toiminnasta eri tilanteissa ilman, että annetaan oikeita vastauksia. (Merilehto 2018, 19.)

Yksittäistä koneoppimisongelmaa tai -menetelmää saattaa olla vaikea lokeroida, koska eri kategorioiden väliset rajat ovat jokseenkin epämääräisiä. Esimerkiksi niin sanottu puoliohjatun koneoppiminen sijoittuu ohjatun ja ohjaamattoman koneoppimisen välimaastoon. (Helsingin yliopisto & Reaktor Innovations 2021.) Yleisimmin käytetyt algoritmit käyttävät regressiota ja luokitusta ennustamaan kohdekattegorioita, etsimään epätavallisia datapisteitä, ennustamaan arvoja ja löytämään yhtäläisyyksiä (Microsoft 2021c).

Algoritmit voidaan jakaa kolmeen eri tyyppiin seuraavasti:

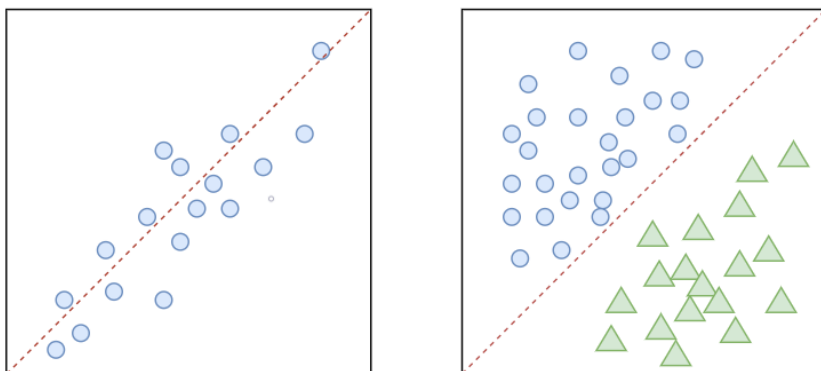
- Regressioalgoritmit, joita käytetään ennustamaan yhtä tai useampaa jatkuvaa muuttujaa perustuen tietojoukon muihin ominaisuuksiin, kuten voitto tai tappio.
- Luokitusalgoritmit, joita käytetään luokittelemaan tiedot eri luokkiin. Luokkia voidaan sitten käyttää ennustamaan yhtä tai useampaa diskreettiä eli epäjatkovaa muuttujaa tietojoukon muiden ominaisuuksien perusteella.

- Klusterointialgoritmit, jotka määrittävät tietojoukon luonnolliset ryhmittelyt ja mallit, ja niitä käytetään ennustamaan tietyn muuttujan ryhmittelyluokituksia. (Barnes 2015, 27.) Klusteroinnissa eli ryhmittelyssä erona luokitteluun on se, että koneelle annetaan dataa, jota ei ole valmiiksi luokiteltu.

2.2.1 Ohjattu oppiminen

Ohjattu oppiminen on usein käytetty ja hyödyllinen koneoppimisen muoto käytännön sovelluksissa. Se on ihmisten ohjaamaa oppimista. Ohjattua oppimista voidaan verrata koulun opetussuunnitelmaan, jossa ennalta tiedetään, mitä oppilaan on koulun päättyessä osattava (Collin & Saarelainen 2016, 211). Järjestelmällä on niin sanottu ystävällismielinen opettaja, joka kertoo mitä järjestelmän tulisi tehdä kussakin tilanteessa. Ohjatussa oppimisessa koneelle annetaan syöte ja oikea ratkaisu opetusdatasta. Tavoitteena on saavuttaa haluttu lopputulos tietynlaisella syötteellä opetusdatan sekä algoritmin kehittämisen avulla. Algoritmilta opetetaan opetusdatalla syöte - tulos parien avulla halutunlaista käytöstä. Toisin sanoen, kun opetusdatasta tiedetään sekä syöte että tulos, tutkimusdatan syötteistä ei tiedetä tulosta. Opetusdatan avulla luodaan luokittimia, joiden avulla jokainen tutkimusdatan syöte voidaan yleistää yhteen luokittimeen tuloksen saamiseksi. (Kokkarainen & Ala-Mutka 2002, 315–316; Kapitanova & Son 2012, 5.)

Ohjattu oppiminen voidaan edelleen jakaa regressio- ja luokittelualgoritmeihin. Regressiota ja luokittelua on havainnollistettu kuvassa 3. Regressioalgoritmit kouluttavat ja voivat ennustaa yhden tai useamman jatkuvan muuttujan, esimerkiksi voiton tai tappion. Luokittelualgoritmeja käytetään ennakoimaan vastauksia, joissa voi olla vain muutama tiedossa oleva arvo. Luokittelussa yritetään löytää sopiva luokkatunniste, kuten analysoidaan positiiviset tai negatiiviset mielipiteet tai mies- ja naispuoliset henkilöt. (Barnes 2015, 28; Tutorialspoint 2021.)



Kuva 3. Regressio ja luokittelu (mukailtu eCloudvalley 2021)

Jokainen datapiste on merkitty tai liitetty kiinnostavaan luokkaan tai arvoon. Esimerkki luokkaan merkitsemisestä on kuvan määrittäminen joko "kissaksi" tai "koiraksi" tai ympyröiksi tai kolmioiksi kuten kuvassa 3. Arvon merkitsemisestä voidaan käyttää esimerkkinä käytettyyn autoon liittyvää myyntihintaa. Ohjatun oppimisen tavoitteena on tutkia monia tällaisia merkittyjä esimerkkejä ja sitten pystyä tekemään ennusteita tulevista datapisteistä. Esimerkiksi uusien valokuvien liittäminen oikeisiin eläinluokkiin tai tarkan myyntihinnan määrittäminen muille käytetyille autoille. (Microsoft 2021d.) Luokittelussa havaitaan siis syöte, kuten kuva ja yritetään päätellä sen luokka, esimerkiksi käsin kirjoitettujen numeroiden tunnistus (syöte on kuva ja luokka on numero välillä 0, 1, ..., 9) (Helsingin yliopisto & Reaktor Innovations 2021).

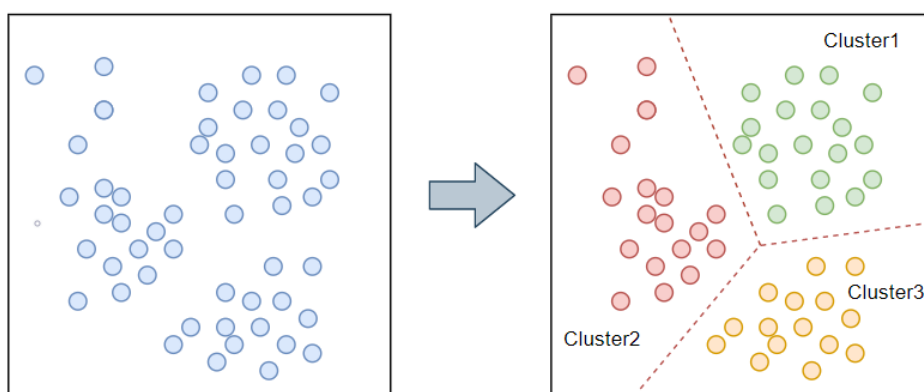
Yleisiä esimerkkejä ohjatusta oppimisesta ovat sähköpostien roskapostisuodatus, verkkosivujen luokittelu niiden sisällön perusteella, äänentunnistus sekä suoratoistopalveluiden ja verkkokauppojen antamat suositukset aiemman toiminnan perusteella (Tutorialspoint 2021). Teollisuudessa menetelmää voidaan esimerkiksi käyttää ruokkimaan kehitettävää algoritmia joukolla tunnettuihin vikaantumisiin liittyvää anturidataa ja normaalia toimintaa sisältävää dataa. Tarkoituksena on, että algoritmi oppii erottamaan huonon ja hyvän toisistaan. Tämän seurauksena syntyvä ennustemalli osaa ennalta hälyttää uhkaavasta vikaantumisesta, kun opitun kaltaista dataa virtaa tietovarastoon. (Collin & Saarelainen 2016, 211.)

2.2.2 Ohjaamaton oppiminen

Ohjaamattomassa oppimisessä kone ei saa valmiita oikeita vastauksia vaan se päättelee ne datassa olevien säännönmukaisuuksien perusteella (itewiki 2021). Tarkoituksena on sitten luoda algoritmi, joka etsii suuresta datamäärästä rakenteita ilman, että datasta tiedetään mitään ennalta. Tavoitteena on etsiä piilossa olevia riippuvuuksia datasta. Ohjaamaton oppiminen on siis ohjatun oppimisen vastakohta ja sitä voidaan pitää menetelmistä vaikeimpana. Ohjaamattomassa oppimisessä algoritmin annetaan päätellä ilman ulkopuolista avustusta. Ohjaamaton oppiminen ei myöskään käytä syöte-tulos-pareja. Uuden ennustemallin menestys riippuu täysin algoritmin kyvystä päätellä ja tunnistaa tulevan tietojoukon malleja, rakenteita ja suhteita. Algoritmin valinnalla on myös suurempi merkitys kuin ohjatussa oppimisessä. (Kapitanova & Son 2012, 17; Barnes 2015, 33–34; Collin & Saarelainen 2016, 211; Lee 2019, 5–7.)

Ohjaamattoman oppimisen tavoite on järjestää tiedot jollakin tavalla tai kuvata niiden rakennetta. Ohjaamaton koneoppiminen voi myös tarkoittaa klusterointia eli ryvästämistä (Kuva 4). Tavoitteena on muodostaa ryhmiä, joihin kuuluvat ovat keskenään samankaltaisia, mutta eroavat muihin ryhmiin kuuluvista. (Helsingin yliopisto & Reaktor Innovations 2021.)

Ohjaamattoman oppimisen avulla voidaan esimerkiksi ryhmitellä ostotavoiltaan samanlaiset asiakkaat asiakastietojen perusteella ja hyödyntää saatuja tuloksia markkinoinnissa. Käytettyjä klusterointialgoritmeja ovat muun muassa K-means klusterointi, itseorganisoituvat kartat (SOM) ja adaptiivisen resonanssin teoria (ART). (Kapitanova & Son 2012, 17; Microsoft 2021c.) Teollisuussovelluksissa ohjaamatonta oppimista käytetään esimerkiksi tunnistamaan anomaliaita, jotka voivat olla varhaisia vikaantumisen oireita (Collin & Saarelainen 2016, 211).

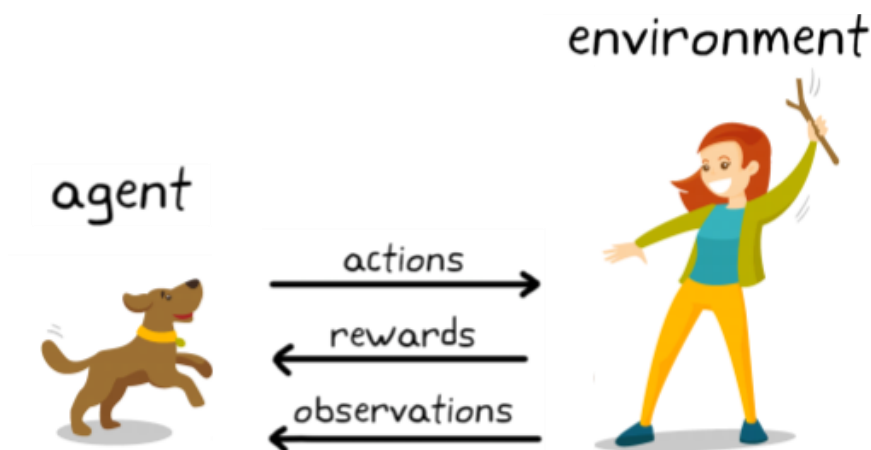


Kuva 4. Klusterointi eli ryvästys (mukailtu eCloudvalley 2021)

2.2.3 Vahvistusoppiminen

Vahvistusoppimisessa oppimista ohjaa ympäristön palaute, ja oppimistekniikka on iteratiivinen ja mukautuva. Sen uskotaan olevan lähempänä ihmisen oppimista. (Gollapudi & Laxmikanth 2016, 343.) Vahvistusoppiminen on vuorovaikutusta. Vahvistusoppimisessa pyritään ohjaamaan algoritmin toimintaa palkitsemalla ratkaisuja, jotka ovat lähellä haluttuja tuloksia, jolloin algoritmi kehittyy askeleittain kohti tavoiteltua tulosta. Algoritmi saa hyvästä ratkaisusta myönteistä palautetta. Mitä parempi ratkaisu on, sitä runsaampaa kiitosta se saa. Huonoista ratkaisuista palaute on samalla tavalla kielteistä. Ehdotonta oikeaa ja väärää ei ole. Järjestelmä arvioi suorituskyykyään palautteen perusteella ja reagoi vastaavasti. Algoritmi rakentaa itselleen politiikan, joka ohjaa toimintaa. Tekniikka soveltuu hyvin käytettäväksi automaattisissa järjestelmissä, joiden on tehtävä paljon pieniä päätöksiä ilman ihmisen ohjausta. (Collin & Saarelainen 2016, 211; Gollapudi & Laxmikanth 2016, 22–23; Microsoft 2021c.)

Kuvassa 5 on havainnollistettu koiran kouluttamisesimerkillä, kuinka vahvistusoppimisalgoritmi toimii. Agentti (*agent*) havainnoi (*observations*) ympäristön tilaa (*environment*) ja toimii (*actions*) sen perusteella. Tämän jälkeen ympäristö eli kouluttaja lähettää agentille eli koiralle toiminnosta palkkion (*reward*) sekä uuden tilan, jonka agentti havainnoi ja valitsee uuden toiminnon.



Kuva 5. Vahvistusoppiminen (Gupta 2019)

Vahvistusoppimista käytetään usein tilanteissa, joissa tekoälyagentin, kuten itseajavan auton, pitää pystyä operoimaan monimutkaisessa ympäristössä ja palaute ratkaisusta tulee mahdollisesti viiveellä. Vahvistusoppiminen käy myös peleihin, joissa lopputulos ratkeaa vasta pelin lopussa. (Helsingin yliopisto & Reaktor Innovations 2021.) Tunnetuimpia esimerkkejä vahvistusoppimisesta ovatkin itseohjautuvat autot ja shakkimestarialgoritmi AlphaGo (Tutorialspoint 2021). Kun esimerkiksi suunnitellaan itseohjautuvaa autoa, halutaan varmistaa, että se noudattaa lakia ja ei ole vaaraksi ihmisille. Auton kokemuksen ja vahvistushistorian myötä se oppii pysymään kaistallaan, noudattamaan nopeusrajoituksia ja huomioimaan jalankulkijat. (Microsoft 2021c.)

Vahvistusoppimisessa algoritmi saa valita toiminnon kunkin datapisteen perusteella. Robottiikassa tämä on yleinen lähestymistapa, kun anturin lukujen joukko muodostaa yhdessä vaiheessa datapisteen ja algoritmin on valittava robotin seuraava toiminto. Se sopii myös luonnollisesti esineiden internet -sovelluksiin. Oppimisalgoritmi saa palautesignaalin onnistumisesta lyhyessä ajassa, mikä osoittaa, kuinka hyvä päätös oli. Tämän signaalin perusteella algoritmi muuttaa strategiaansa saavuttaakseen paremman palautteen. (Microsoft 2021d.) Vahvistusoppiminen tulee todennäköisesti vastaan teollisessa internetissä ja älykäden tehtäiden kehittämisessä (Collin & Saarelainen 2016, 211).

2.3 Koneoppimismallin arviointi

Koneoppimisaineisto jaetaan opetus- ja testidataan, jolloin voidaan välttää suuremmat koneoppimismallin koulutuksessa tapahtuvat virheet. Opetusvaiheessa opetusdatalla opetettu malli pystyy ennustamaan vastauksen mille tahansa syötteelle. Vaikka mallin ennustustarkkuus testidatassa antaa hyvän kuvan mallin tarkkuudesta, koneoppimisen tuloksena saatavan mallin ennustustarkkuus voi kuitenkin olla hyvin erilainen opetusdataa ja testidataa käytettäessä. Tämä johtuu niin sanotusta ylisovittamisesta. (Helsingin yliopisto & Reaktor Innovations 2021).

Ylisovittaminen eli over-fitting tarkoittaa toisin sanoen sitä, että malli sopii dataan liian hyvin. Opetusdata ei edusta tällöin kattavasti laajempaa havaintopopulaatiota. Syynä voi olla datassa oleva satunnaiskohina tai aineiston pieni koko. Lisäksi eri luokkamuuttujien tasojen välisten kombinaatioiden suuri määrä pienentää efektiivistä otoskokoja. Ennustemallia tehtäessä data jaetaan kahteen tai useampaan osaan yleensä satunnaisotannalla. Suurin osa datasta käytetään opetusdataksi ja loput testidataksi. Tässä työssä satunnaisesti jaetusta datasta opetusdatan osuus oli 70 % ja testidatan 30 %. Testidatalla validoinnin lisäksi suositellaan opetusdatan pilkkomista sisäisesti useaan eri opetus- ja testidataan, joilla suoritetaan mallin opetus ja validointi. Menetelmää kutsutaan ristiin validoinniksi. (Ruokolainen 2018.) Koneoppimismallien arviointiin voidaan käyttää useita eri mittareita. Tässä opinnäytetyössä käytetyistä mittareista kerrotaan tarkemmin luvussa 5.6.

2.4 Koneoppiminen pilvipalveluna

Pilvipalvelu on IT-palvelun toimittamista internetin eli pilven kautta. Se tarjoaa käyttäjilleen tarpeen mukaan, skaalautuvia, laitteista riippumattomia palveluja. Palvelusta maksetaan käytön mukaan, se on nopeasti saatavilla ja sen teho määräytyy tarpeen mukaan. (Harkut 2019.) Tämä tarkoittaa siis tietoteknisen infrastruktuurin tai datakeskuksien omistamisen sijaan vuokraamista pilvipalvelun tarjoajalta. Palvelut käsittävät tallennus, verkko- ja prosessointitehosta luonnollisen kielen käsittelyyn ja tekoälyyn sekä tavallisiin toimistosovelluksiin. Pilvipalvelu on terminä ollut käytössä 2000-luvun alkupuolelta, mutta tietojenkäsittely palveluna on ollut olemassa jo 1960-luvulla. Tietokonetoimistot antoivat silloin yritysten vuokrata aikaa keskustietokoneita sen sijaan, että joutuivat ostamaan koneen itse. (Ranger 2018.)

Yleisemmin pilvipalvelut on jaettu kolmeen palvelumalliin eli seuraaviin palveluihin: SaaS, IaaS ja PaaS. Pilvipalveluiden oikean palvelumallin tai palvelumallien yhdistelmän valitseminen on kriittinen menestystekijä. (Kavis 2014, 13.) SaaS tarkoittaa ohjelmiston toimittamista palveluna, IaaS infrastruktuuria palveluna ja PaaS alustaa palveluna. Koneoppiminen

palveluna eli MLaaS taas viittaa useisiin palveluihin, jotka tarjoavat koneoppimisen työkaluja pilvipalvelujen osana. MLaaS-palveluntarjoajat tarjoavat kehittäjille palveluja, jotka sisältävät ennakoivan analyysin, tiedonsiirrot ja visualisoinnit, datan mallinnus-API:t, kasvotunnistuksen, luonnollisen kielen prosessoinnin ja koneen syväoppimisalgoritmit. (Wolhuter 2021.)

Kuznetsovin (2018, 2) mukaan koneoppiminen palveluna tarkoittaa palveluita kuten datan esiprosessoinnin, mallien opettamisen ja arvioinnin sekä niiden avulla ennustamisen, joita pääsee normaalisti käyttämään REST API:n kautta. MLaaS-ratkaisujen etuna on niiden käyttöönoton ja käyttämisen helppous. Toisaalta usein ennakoon määritellyt mallit ovat hyvin yleistettyjä eivätkä siten tarpeeksi erikoistuneita kyseessä olevaan ongelmaan. MLaaS-palveluita ovat esimerkiksi Microsoft Azure Machine Learning, Amazon ML, Google Cloud AI ja IBM Watson. Tässä työssä tutustutaan tarkemmin Microsoft Azuren Machine Learning -palveluun.

3 Koneoppimisen sovelluskohteet

3.1 Yritysten sovelluskohteet

Tekoäly ja koneoppimisen sovellukset ovat tulleet tutuiksi useissa kuluttajasovelluksissa kuten edellisessä luvussa on mainittu. Koneoppimisen käytöllä on tutkimuksissa havaittu olevan suurta hyötyä yrityksille kannattavuuden paranemisena, kun toimintaa on pystytty tehostamaan, tulot ovat kasvaneet ja käyttökustannukset alentuneet (Pardeshi 2021). Koneoppimisen avulla pystytään kaupallisessa käytössä luomaan monimutkaisia algoritmeja ja malleja. Niitä voidaan hyödyntää esimerkiksi päätöksenteon tukena tai prosessien tehostamisessa. Koneoppimisen lukuisia sovelluskohteita löytyykin muun muassa hakukoneissa, lääketieteessä, robotiikassa ja tietokonepeleissä. (itewiki 2021.) Teollisuudessa koneoppiminen on usein menetelmä teollisen internetin datan louhinnassa, koska sillä voidaan luoda ennustemalleja esimerkiksi laitteen vikaantumisesta (Collin & Saarelainen 2016, 211).

Teollisuuden evoluutiossa on siirrytty vaiheeseen Teollisuus 4.0, joka on tunnetuin strateginen muutosohjelma teollisen internetin mahdollisuuksien hyödyntämiseksi. Se on alun perin Saksan hallituksen vuodesta 2012 lanseeraama hanke maan valmistavan teollisuuden kilpailukyvyyn säilyttämiseksi ja vahvistamiseksi. Perustana on vahva osaaminen automaatioissa, toiminnanohjauksessa ja sulautetuissa järjestelmissä. (Collin & Saarelainen 2016, 37.)

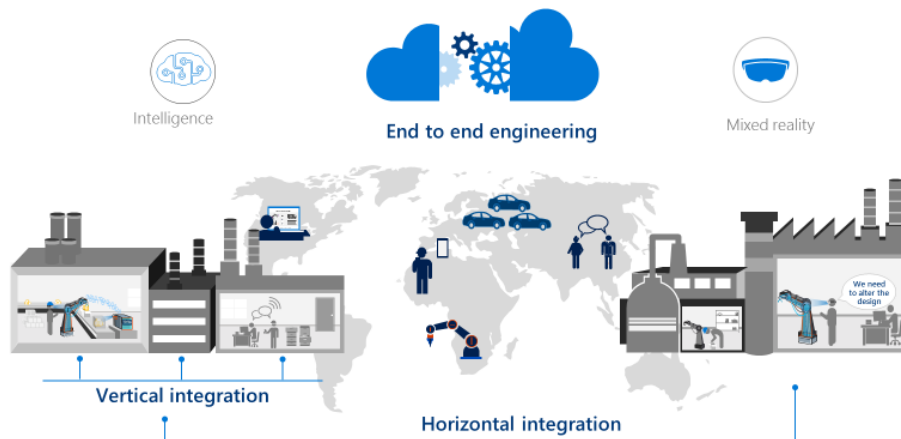
3.2 Esineiden internet

Esineiden internetillä (IoT) on paljon tarjottavaa eri aloille, erityisesti teollisuudelle kuten laitteiden ennakoivalle kunnossapidolle tai laadun valvonnassa. Nämä ovat keskeisiä aiheita myös opetuksessa varsinkin tekniikan alalla. IoT ja koneoppiminen toimivat hyvin yhdessä. IoT tuottaa valtavan määrän numeerista dataa, jota voidaan koneoppimisen avulla hyödyntää IoT-ratkaisujen kehittämiseen.

Meldrumin (2017) mukaan tulevaisuuden tehtaot ovat monimutkaisia valmistusekosysteemejä käsittäen itsesäätäviä koneita ja tehtaita, jotka voivat mukauttaa tuotantoa. Niissä on mukautuva avoin arvoketju, joka lisää kykyä valmistaa yksilöllisiä tuotteita lyhyellä elinkaarella. Teollisuus 4.0:n seurauksena rajat digitaalisten ja fyysisten toimintojen välillä ovat hämärtyneet. Tulevaisuuden tehtaassa (Kuva 6) tämä tapahtuu sekä horisontaalisesti organisaatioiden välillä että vertikaalisesti organisaation sisällä muuttaen valmistusekosysteemiä. Tehdas hyödyntää IIoT:ia eli teollista esineiden internetiä, kehittyntä automaatiota, tekoälyä ja tilannetietoisuuden kognitiivista lähestymistapaa. Teollisen IoT:n, yhdistetyn

todellisuuden (*mixed reality*), materiaalia lisäävän valmistuksen ja data-analytiikan ansiosta valmistajat voivat keskittyä tehokkuuden parantamiseen ja tuotannon personointiin.

Key concepts for the Factory of the Future



Kuva 6. Tulevaisuuden tehdas (Meldrum 2017)

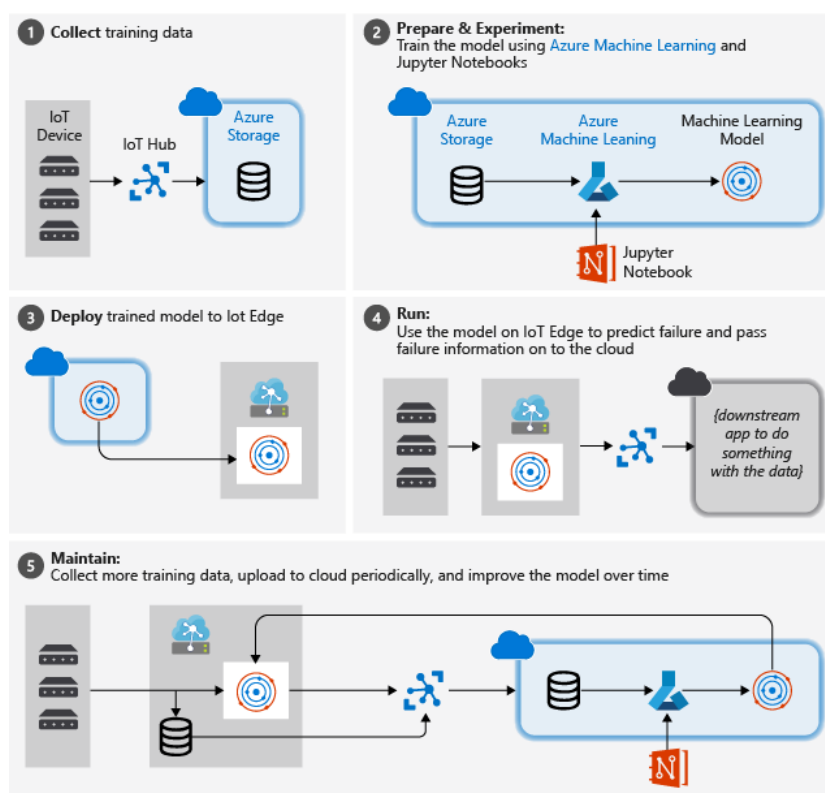
Tulevaisuuden tehtaiden virtuaalisissa valmistusekosysteemeissä ihmiset ja koneet toimivat yhdessä työtovereina ja yhteistyörobotit (kobotit) ovat riittävän älykkäitä antamaan työpanoksensa laajennetun virtuaalisen ekosysteemin välityksellä. Koneoppimisen, ennakoivan analytiikan ja pilvipohjaisen datan perusteella valmistajat tietävät etukäteen, milloin laitteeseen saattaa tulla vika sen sijaan, että vika toteutuu ja käsitellään seurauksia ja seisokkeja. Tämän ansiosta saanto kasvaa, hävikki ja lisätyön määrä vähenee ja valmistusjakson kesto lyhenee. Optimoinnin ja oivallusten myötä nykyaikaiset valmistajat voivat toimia asiakaskeskeisesti, innovoida nopeammin ja tulla ketterämmäksi. Azure-alusta ja Azure Machine Learning auttavat hyperkytkeytynyttä maailmaa toimimaan älykkäiden pilvimallien kanssa. Tämä luo perustan paitsi esineiden internetille myös esineiden, ihmisten ja palveluiden internetille, jotka toimivat yhdessä tulevaisuuden tehtaissa. (Meldrum 2017.)

Koneilta ja laitteilta saatava sensoridatan pilveen tallentamisen etuina ovat erittäin edullinen tallennustila ja automaattinen skaalautuminen miljoonien laitteiden datavirroille. Tällaisissa pilvipalveluissa on pelkän datavaraston lisäksi laitehallinta, analytiikka ja visualisoinnin työkalut. Teollisen internetin tietovarasto voi käsittää myös datamassat, kuten toiminnan- tai tuotannonohjausjärjestelmän datan. Niiden datavirrat on mahdollista integroida rajapintojen kautta varsinaiseen teollisen internetin tietovarastoon, joka voidaan nähdä osin virtuaalisena lisäkerroksena monen eri tietolähteen päällä. (Collin & Saarelainen 2016, 202–203.)

3.3 IoT ja Azure Machine Learning

IoT-sovelluksissa voidaan hyödyntää älykästä pilveä ja älykästä ulkoreunaa (Microsoft 2019). Älykkäällä pilvellä tarkoitetaan usein tekoälyn, kuten analytiikan, koneoppimisen ja kognitiivisten palvelujen, hyödyntämistä ja soveltamista erilaisissa käyttökohteissa. Pilven reunalla tapahtuvat palveluiden soveltaminen ja saatava hyöty, toisin sanoen lähellä varsinaista käyttökohdetta, joka saattaa olla toiminta- tai aikakriittinen. Microsoft käyttää älykkästä pilven reunasta termiä Intelligent Edge, josta voidaan käyttää suomeksi esimerkiksi nimitystä älykäs ulkoreuna. Termillä kuvataan kaikkia laitteita, jotka laajentavat pilven lähemmäksi sovelluskohdetta. Pilven reunalla laitteet vaihtelevat aina IoT-sovelluksista, virtuaalisen-, lisätyn- tai tehostetun todellisuuden laitteista, pilveä laajentaviin loppukäyttäjän tai palvelinkeskuksen laitteisiin. (Hypönen 2018.)

Azuresa kokonaisratkaisu voidaan toteuttaa Azure Machine Learningin ja IoT Edgen avulla. Toimivan mallin luominen ja sen käyttäminen IoT-datan käsittelyyn sisältää koneoppimismallin kouluttamisen pilvipalvelun IoT-laitteista kerätyn datan avulla, mallin käyttöönoton IoT Edge -palvelussa sekä mallin ylläpidon ja hienosäädön. Kuvassa 7 on prosessi toimivan mallin luomisesta ja sen käyttämisestä IoT-datan käsittelyssä. (Microsoft 2019.)



Kuva 7. Azuren koneoppimismalli IoT-sovelluksessa (Microsoft 2019)

Prosessi etenee seuraavien vaiheiden mukaisesti:

1. Harjoitusdatan kerääminen. Prosessi alkaa harjoitusdatan keräämisellä. Joissakin tapauksissa dataa on jo kerättynä ja se on saatavilla tietokannassa tai datatiedostoina. Muissa tapauksissa, erityisesti IoT-skenaarioissa, tiedot on kerättävä IoT -laitteista ja antureista ja tallennettava pilveen.
2. Datan valmistelu. Useimmissa tapauksissa laitteista ja antureista kerätty raakadata vaatii valmistelua koneoppimista varten. Tämä vaihe voi sisältää datan puhdistamista, datan uudelleenmuotoilua tai esikäsittelyä.
3. Koneoppimismallin luominen. Valmistellun datan pohjalta kokeillaan erilaisia koneoppimisalgoritmeja ja parametreja mallien kouluttamiseksi ja tulosten vertaamiseksi toisiinsa.
4. Mallin käyttöönotto. Kun malli täyttää asetetut onnistumiskriteerit, voidaan siirtyä käyttöönottoon. Tämä edellyttää mallin paketoimista verkkopalvelusovellukseen, johon voidaan syöttää dataa käyttäen REST-rajapintakutsuja ja analyysituloksia. Verkkopalvelusovellus pakataan sitten Docker-konttiin, joka puolestaan voidaan ottaa käyttöön joko pilvessä tai IoT Edge -moduulina.
5. Mallin ylläpito ja jalostaminen. Työ ei pääty mallin käyttöönoton jälkeenkään. Useissa tapauksissa jatkuu datan kerääminen ja sen lataaminen säännöllisesti pilveen. Dataa voidaan käyttää uudelleen kouluttamiseen ja parantamaan mallia, joka voidaan ottaa sitten uudelleen käyttöön IoT Edgessä. (Microsoft 2019.)

4 Azure Machine Learning

4.1 Ominaisuudet

Azure Machine Learning on Microsoft Azuren tarjoama pilvipohjainen ympäristö koneoppimismallien kouluttamiseen, käyttöönottoon, automatisointiin, hallintaan ja seurantaan. Azure-koneoppimista voidaan käyttää kaikenlaiseen koneoppimiseen, klassisesta ML:stä syväoppimiseen, ohjattuun ja ohjaamattomaan oppimiseen. Azure Machine Learning -työtilassa (Workspace) voi rakentaa, kouluttaa ja seurata koneoppimisen ja syväoppimisen malleja kirjoittamalla Python- tai R-koodin kehitystyökalujen avulla tai Azure Machine Learning Studiossa. Palvelu on myös yhteensopiva suosittujen syväoppimisen ja avoimen lähdekoodin käyttöä tukevien työkalujen kanssa, kuten PyTorch, TensorFlow, scikit-learn ja Ray RLLib. (Microsoft 2021b.) Azure Machine Learning soveltuu eri kohderyhmille niin datatieteilijöille, kehittäjille kuin data-analytiikoillekin (Baker ym. 2021).

Azure Machine Learningiä voidaan käyttää ainakin neljään eri tarkoitukseen:

1. Arvojen ennustaminen. Regressioalgoritmit auttavat muuttujien välisten syiden ja seurausten tunnistamisessa luomalla arvoista mallin, jota sitten käytetään ennusteen tekemiseen. Regressiotutkimuksien avulla voidaan ennakoida tulevaisuutta ja siten esimerkiksi ennakoida kysyntää, ennustamaan myyntilukuja tai arvioimaan kampanjan tuloksia.
2. Epätavallisten tapahtumien tunnistaminen. Koneoppimista käytetään mahdollisen riskin havaitsemiseen, kun poikkeavuuksien havaitsemisalgoritmit osoittavat datan, joka on ennakoidun normin ulkopuolella. Laitteiden toimintahäiriöt, rakenteelliset viat, tekstivirheet ja petostapaukset ovat esimerkkejä, joiden tunnistamiseen ja niistä aiheutuvien huolenaiheiden ratkaisemiseen koneoppimista käytetään.
3. Rakenteen etsiminen. Klusterointialgoritmit paljastavat usein ensimmäisenä askeleena koneoppimisessa tietojoukon taustalla olevan rakenteen. Luokittelussa, klusterointia käytetään esimerkiksi markkinoiden segmentoinnissa hinnoittelun ja asiakkaiden mieltymysten ennakoimiseksi.
4. Luokkien ennustaminen. Luokitusalgoritmit auttavat määrittämään oikean luokan datalle. Luokittelua käytetään ohjatussa oppimisessa, jossa annetaan ennalta määritellyt tunnisteet. (Microsoft 2021a.)

4.2 Azure Machine Learning Studio

Azure Machine Learning -studio on Azure Machine Learning -verkkoportaali, joka tarjoaa matalan koodin ja koodittomia vaihtoehtoja mallin kouluttamiseen, käyttöönottoon ja hallintaan. Studio integroituu Azure Machine Learningin kehitystyökalujen kanssa kuten Python SDK, Azure Resource Manager REST API ja CLI v2. Integroinnit Azure-palveluihin tukevat koneoppimisprojektia kokonaisuudessaan. Näitä ovat tallennus- ja tietokanta -vaihtoehdot kuten Azure SQL Database ja Azure Storage Blobs. (Microsoft 2021b.)

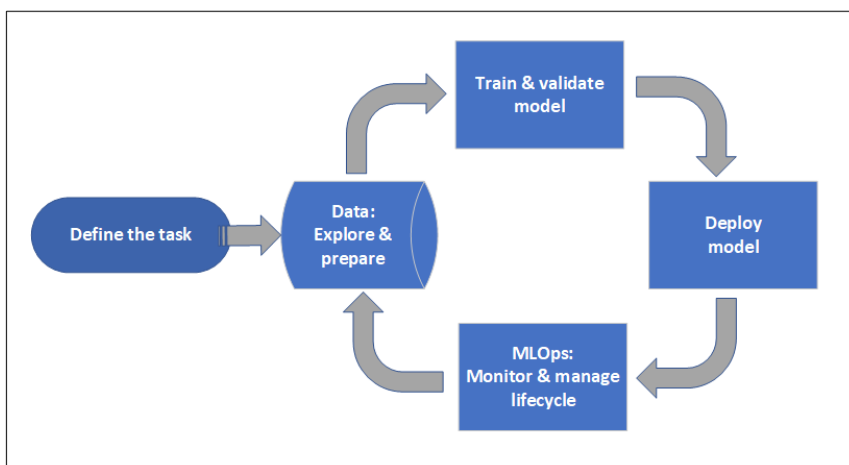
Azure Machine Learning Studio on graafinen käyttöliittymä projektityötilalle. Studiossa voidaan

- tarkastella ajoja, mittareita, lokeja ja tuloksia
- kirjoittaa ja muokata koodia ja tiedostoja
- hallita datan kirjautumistietoja, laskentaa ja ympäristöjä
- visualisoida mittareita, tuloksia ja raportteja tai kehitysrajapintojen kautta kirjoitettuja putkia (pipelines)
- kirjoittaa AutoML-tehtäviä.

Lisäksi Designer-suunnittelutyökalulla on vedä ja pudota -käyttöliittymä, jonka avulla voidaan kouluttaa ja ottaa käyttöön malleja. (Microsoft 2021b.) Designer tarjoaa visuaalisen suunnittelutyökalun, jolla voidaan muun muassa yhdistää datasettejä ja -moduuleja putkien ja ML-mallien luomiseksi, testaamiseksi, kouluttamiseksi ja käyttöönottamiseksi (Baker ym. 2021).

4.3 Koneoppimisprojektin työnkulku

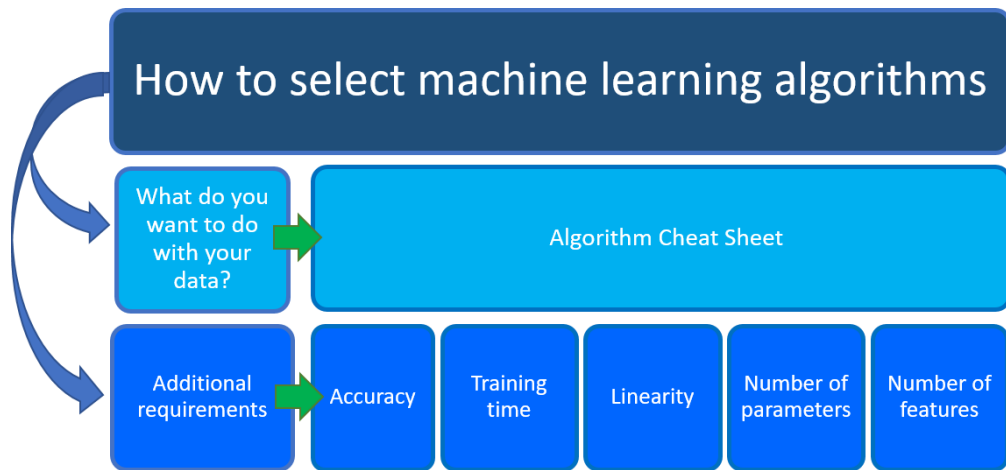
Koneoppimismalleja kehitetään tyypillisesti osana hanketta, jolla on tavoite ja päämäärät. Kun kokeillaan dataa, algoritmeja ja malleja, kehitys on iteratiivista. Edellisessä luvussa kuvattiin jo projektin kulkua IoT-sovelluksessa (Kuva 7). Vaikka projektin elinkaari voi vaihdella projektikohtaisesti, se usein näyttää yksinkertaistettuna kuten on esitetty kuvassa 8. (Microsoft 2021b.)



Kuva 8. Projektin elinkaari (Microsoft 2021b)

Ensimmäisenä kuvassa on tehtävän määrittely (*Define the task*). Siinä etsitään vastausta kysymyksiin mitä datalla halutaan tehdä ja mihin liiketoiminnan kysymykseen halutaan vastata oppimalla aiemmasta datasta. Lisäksi on selvitettävä vaatimukset kuten mikä on tarkkuus, harjoitusaika, lineaarisuus, parametrien lukumäärä ja ominaisuudet, joita ratkaisu tukee. Nämä tekijät vaikuttavat datan tutkimiseen ja valmisteluun (*Data: Explore & prepare*) sekä koneoppimismallin luomiseen ja algoritmin valintaan (*Train & validate model*) Azure Machine Learningissa. (Microsoft 2020; Microsoft 2021b.) Mallin valinnan jälkeen, otetaan se käyttöön (*Deploy model*). Kun projekti on käyttövalmis, käyttäjien työ voidaan automatisoida koneoppimisputkessa ja käynnistää aikataulun tai HTTPS-pyyntöjen perusteella. Projektin elinkaari jatkuu sen seurannalla ja hallinnalla (*MLOps: Monitor & manage lifecycle*). (Microsoft 2021b.)

Kuvassa 9 on esitetty algoritmin valintaan vaikuttavat tekijät. Azure Machine Learningissä on apuna Algorithm Cheat Sheet (Microsoft 2021d, Liite 1) datan käyttötarkoituksen määrittämiseen. Algorithm Cheat Sheetiltä etsitään tehtävä, joka halutaan tehdä ja etsitään algoritmi ennustavaa analytiikkaratkaisua varten. (Microsoft 2020.)



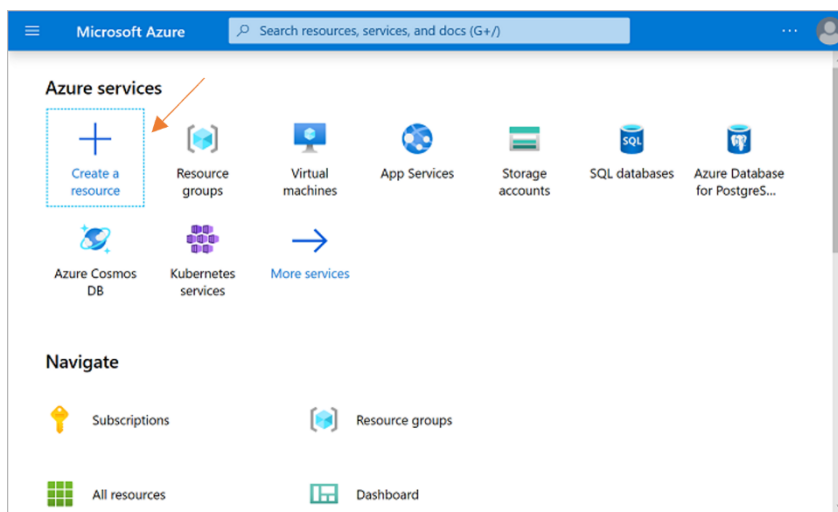
Kuva 9. Algoritmin valinta (Microsoft 2020)

5 Azure Machine Learning käytännössä

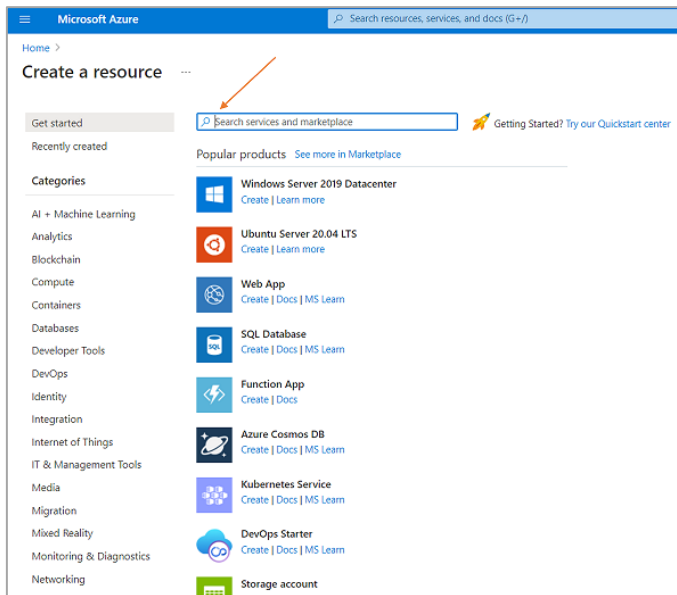
5.1 Palvelun käyttöönotto

Palvelun käyttöönottamiseksi käyttäjällä on oltava Microsoft Azuren tili ja voimassa oleva tilaus. Microsoft tarjoaa Azuren uusille käyttäjille ilmaisen kokeilujakson ja -rahaa. Kokeilujakson jälkeen palvelun käyttöä voi jatkaa edelleen, mutta ilmaisten kuukausierien ylittävältä osuudelta palvelusta laskutetaan (Microsoft 2021e). Vaikka Azure Machine Learningin käytöstä ei peritä lisämaksua, tilaajalta veloitetaan muista kulutetuista Azure -palveluista, kuten laskenta- ja tallennuskustannukset (Microsoft 2021g). Laskentainstanssin hinnat määräytyvät virtuaalikoneen tehon mukaan, tuntihinnat alkaen muutamista senteistä.

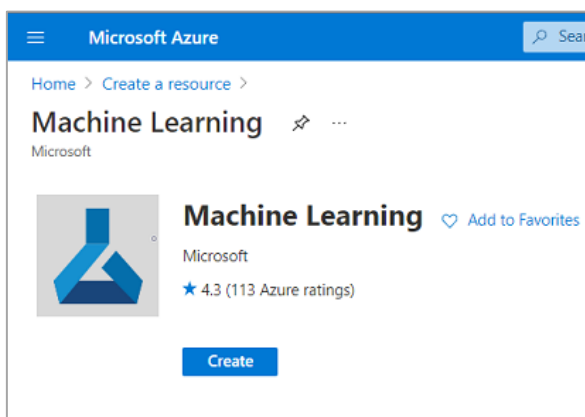
Koneoppimisympäristön luomiseksi Azuren aloitussivulta luodaan uusi resurssi kohdasta Create a resource (Kuva 10), seuraavaksi kirjoitetaan hakukenttään Machine Learning (Kuva 11), jolloin päästään luomaan uusi Machine Learning -työtila valitsemalla Create (Kuva 12).



Kuva 10. Uuden resurssin luominen (mukailtu Microsoft 2021h)



Kuva 11. Machine Learning -resurssin luominen



Kuva 12. Machine Learning -työtilan luominen

Tämän jälkeen määritellään Machine Learning -työtila (Workspace) (Kuva 13). Työtila on koneoppimistoimintojen ylätasoinen resurssi, joka avulla voidaan kouluttaa, ottaa käyttöön, automatisoida, hallita ja seurata koneoppimismalleja (Microsoft 2021f). Ensin täytetään Projektin tiedot Subscription ja Recourse group eli valitaan käytössä oleva tilaus ja käytetään olemassa olevaa resurssiryhmää tai nimetään uusi resurssiryhmä. Seuraavaksi määritellään työtilan tiedoista Workspace name ja Region eli annetaan työtilan yksilöivä nimi ja valitaan lähin alue. Kun kaikki tarvittavat tiedot on syötetty, lopuksi luodaan työtila valitsemalla vasemmasta alakulmasta Review + create ja valitaan vielä Create.

Microsoft Azure Search resources, services, and docs (G+)

Home > Machine learning >

Machine learning

Create a machine learning workspace

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription *

Resource group * [Create new](#)

Workspace details

Specify the name and region for the workspace.

Workspace name * ✓

Region *

Storage account * [Create new](#)

Key vault * [Create new](#)

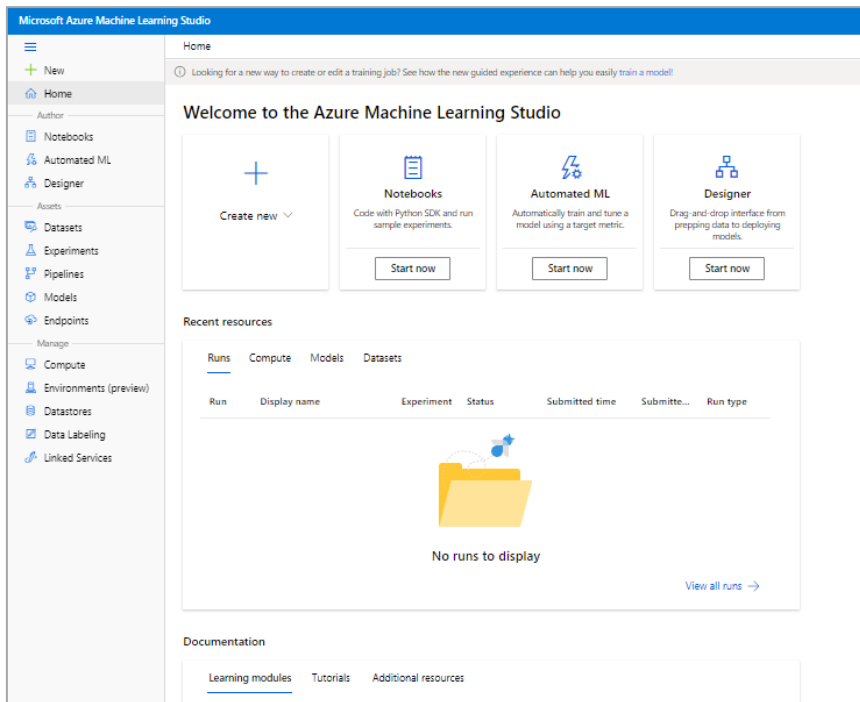
Application insights * [Create new](#)

Container registry * [Create new](#)

[Review + create](#) [< Previous](#) [Next : Networking](#)

Kuva 13. Machine Learning Workspace -työtilan määrittäminen

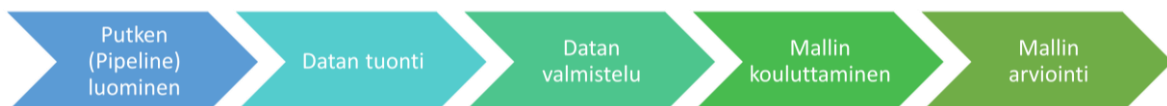
Valmiista työtilasta pääsee avaamaan Azure Machine Learning studio -verkkoportaaliin valitsemalla Launch Studio. Studiossa on mahdollista valita koodillinen (Notebooks), automaattinen (Automated ML) tai täysin kooditon, suunnittelutyökalulla toteutettava (Designer) koneoppimismallin kehittäminen. Azure Machine Learning Studion aloitusnäköymästä (Kuva 14) käyttäjä pääsee kaikkiin studion koneoppimisen palveluihin. Sieltä löytyy myös dokumentointi kuten tutoriaalit.



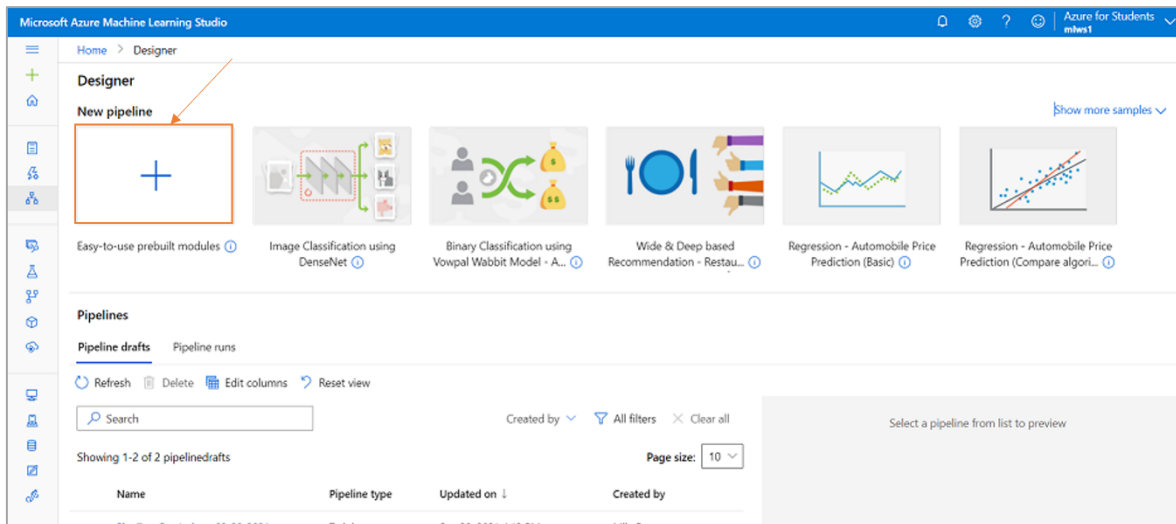
Kuva 14. Azure Machine Learning Studion aloitusnäky

5.2 Mallin kehittäminen suunnittelutyökalulla

Koneoppiminen kehitty nopeasti ja opiskelun aloittamiseksi matalla kynnyksellä tässä työssä käytetään Designer-suunnittelutyökalua. Machine Learning aloitussivulta valitsemalla Designer päästään työkalun aloitusnäkyyn (Kuva 16). Vaihtoehtoina on avata uusi putki, pohja (New pipeline) mallin rakentamiseksi tai käyttää valmiita malleja. Tässä työssä kokeiltiin ensin valmiita regressio- ja luokittelumallipohjia, jonka jälkeen luotiin uusi pohja lämpötilan ennustamiseksi. Koneoppimisprosessi kokonaisuudessaan etenee vaiheittain. Viisi vaihetta koneoppimisen suorittamiseksi on esitetty kuvassa 15.

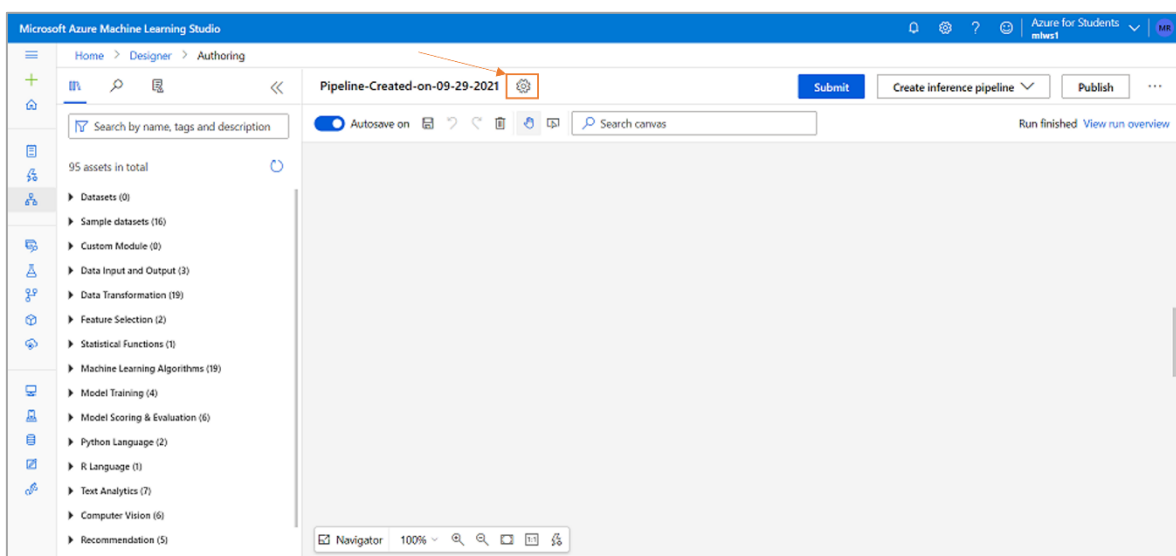


Kuva 15. Koneoppimisprosessi

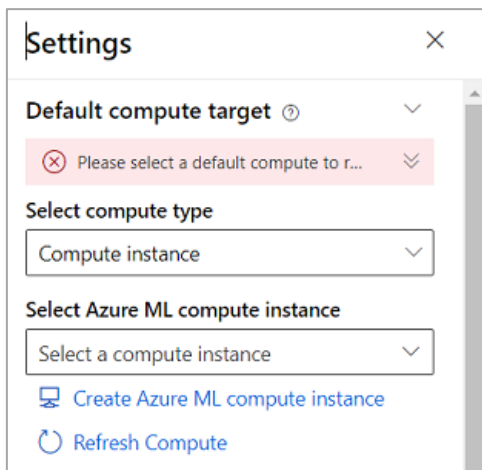


Kuva 16. Designer-aloitusnäkö

Valitsemalla New pipeline (Kuva 16) avautuu uusi, tyhjä pohja (kuva 17). Vasemmalla näkyy lista otsikoita, joista valitsemalla tulee näkyviin useita moduuleita. Niitä voi raahata oikealle kanvaalle (canvas), jonka yläosassa on putken oletusnimi, Pipeline-Created-on. Sen tilalle voidaan vaihtaa uusi nimi, tässä työssä se voi esimerkiksi olla Temperature prediction. Tässä vaiheessa on myös hyvä valita tai luoda uusi laskentainstanssi, johon pääsee asetuksista, ylhäältä ratasikonista (Settings, Kuva 17). Mikäli laskentainstanssia ei ole vielä luotu, valitaan Create Azure ML compute instance (Kuva 18). Muutoin valitaan kohdasta Select Azure ML compute instance haluttu olemassaoleva instanssi.

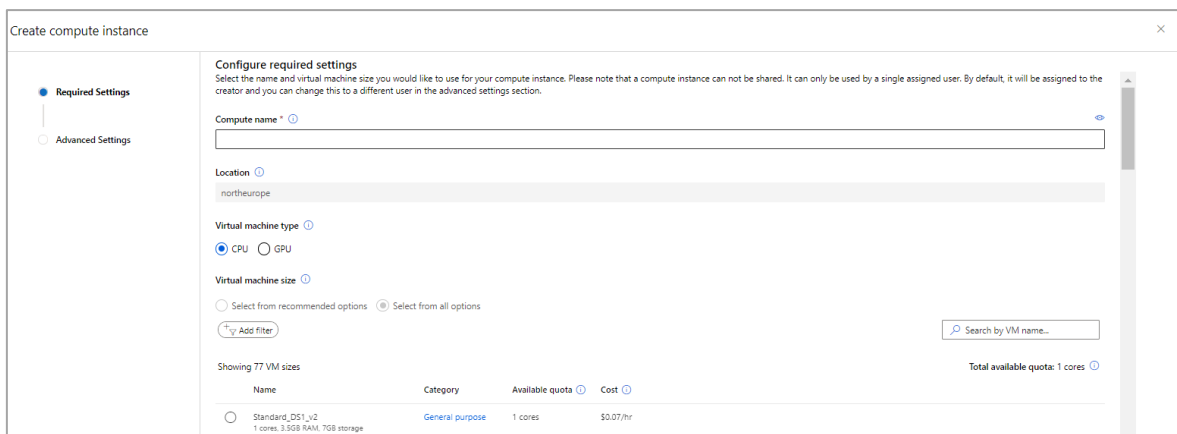


Kuva 17. New pipeline



Kuva 18. Laskentainstanssin luominen

Tehdään vielä laskentainstanssin määrytykset eli annetaan sille nimi (Compute name) ja valitaan listalta sopiva virtuaalikone (Kuva 19).



Kuva 19. Laskentainstanssin määrittelemine

5.3 Datatunnti

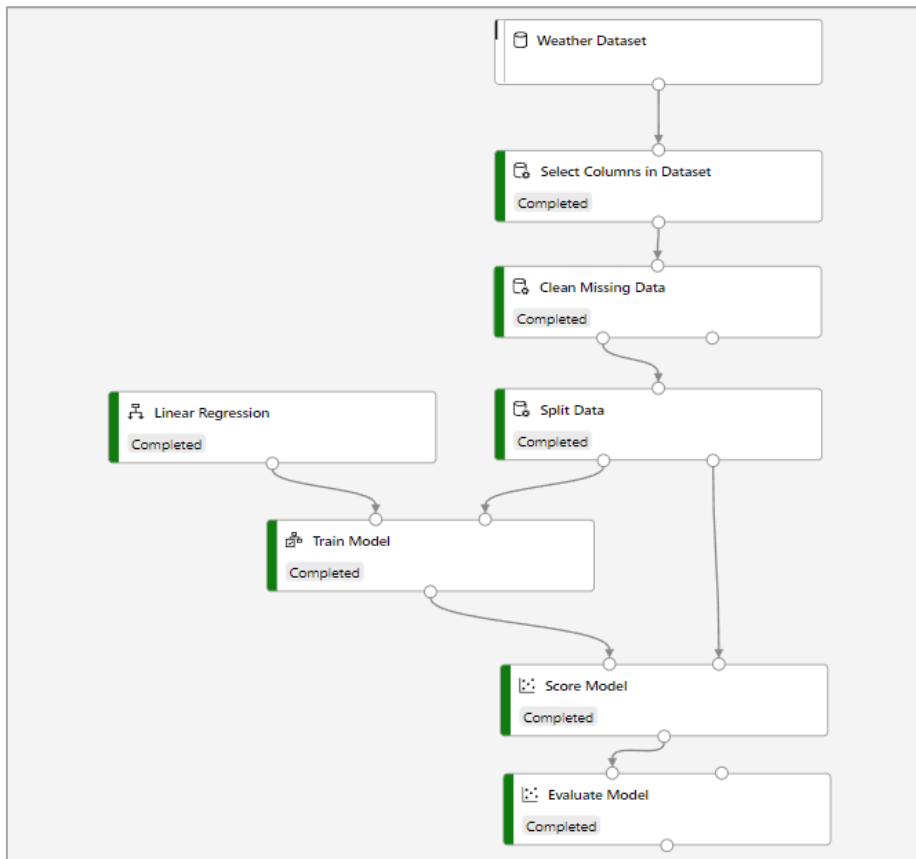
Kanvaan vasemman puolen valikosta löytyvät moduulien lisäksi datasetit (Kuva 17). Raahataan kanvaalle datasetti Datasets tai Sample datasets alta, jos käytetään valmista datasettiä. Datasetin luomiseen voidaan käyttää Azure Storage Blobs -palvelua (Microsoft 2021k).

Tässä työssä harkittiin oman datasetin luomista ja datan hankkimista avoimen datan kautta, mutta lopulta päädyttiin käyttämään erittäin kattavaa valmista, suunnittelutyökalun tarjoamaa Weather-datasettiä. IoT- ja IIoT-sovelluksissa sensoreista saatava data on useimmiten myös numeerisessa muodossa. Koska sovelluskohteista ei ollut saatavilla dataa, valittiin Weather-datasetti ja siitä lämpötila tarkastelun kohteeksi. Datasettiä pääsee tarkastelemaan visuaalisesti valitsemalla hiiren oikealla kanvaalle raahatun datasetin päällä Preview data.

Käytetty datasetti koostui NOAA:n tuntikohtaisista maa-asemien säähavainnoista. Säätiiedot kattavat 70 vilkkaimpien lentoasemien sääasemilla tehdyt havainnot huhti-lokakuussa 2013. Datasetti sisältää 26 saraketta. (Microsoft 2021i.)

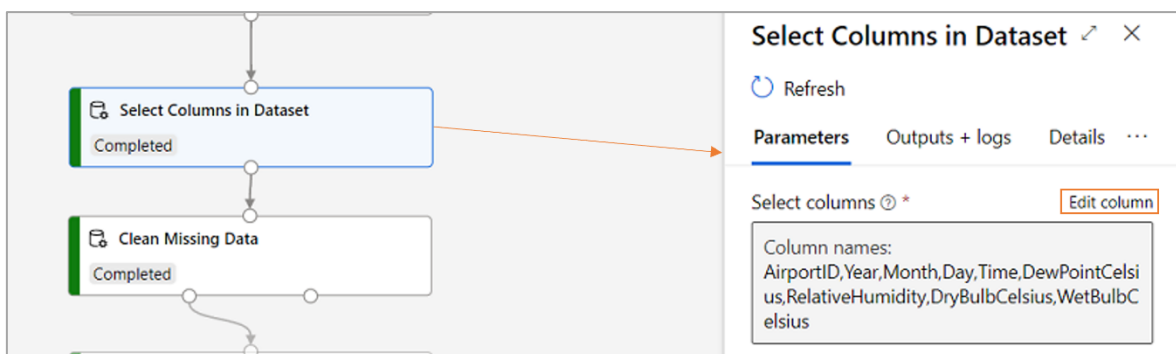
5.4 Datan valmistelu

Datasetit vaativat usein käsittelyä ennen analyysiä. Vain joitain datasetin sarakkeita halutaan ottaa mukaan tai sulkea pois analyysistä. Datasetistä halutaan mahdollisesti poistaa myös tyhjiä rivejä. Tällaiset datan valmistelutehtäviä suorittavat moduulit löytyvät valikosta vasemmalta Data Transformation-otsikon alta. Tässä kokeilussa datan valmistelu tarkoittaa sarakkeiden valitsemisen datasetistä (Select Columns in Dataset) ja datan puhdistamisen (Clean Missing Data), jotka on esitetty valmiissa mallissa (Kuva 20), kuten kaikki muutkin moduulit.

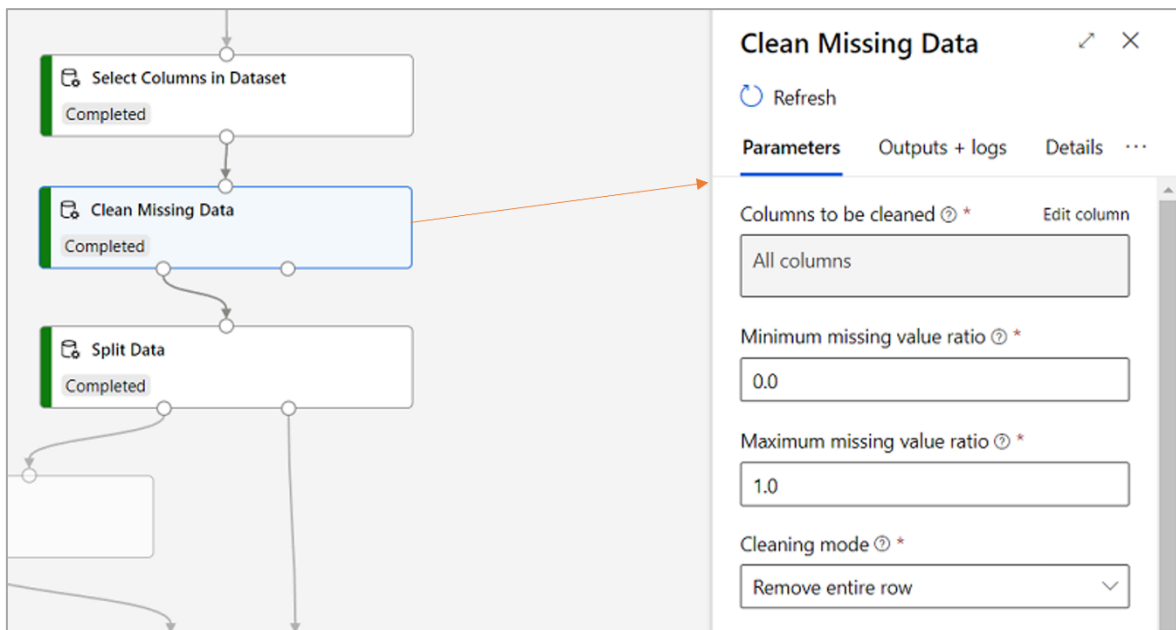


Kuva 20. Lämpötilan ennustaminen regressiomallilla

Tässä kokeilussa mallia varten valittiin ensin mukaan viisi saraketta, joita lisättiin myöhemmin mallin parantamiseksi. Lopulta mukaan otettiin yhdeksän saraketta (Kuva 21). Data-setin sarakkeita pääsee muuttamaan kohdasta Edit column. Jokaisen moduulin tiedoista Parameters alta löytyy Comment-laatikko, johon voi lisätä lyhyen kuvauksen moduulista. Poistetaan vielä tyhjät rivit (Kuva 22).



Kuva 21. Sarakkeiden valitseminen datasetistä

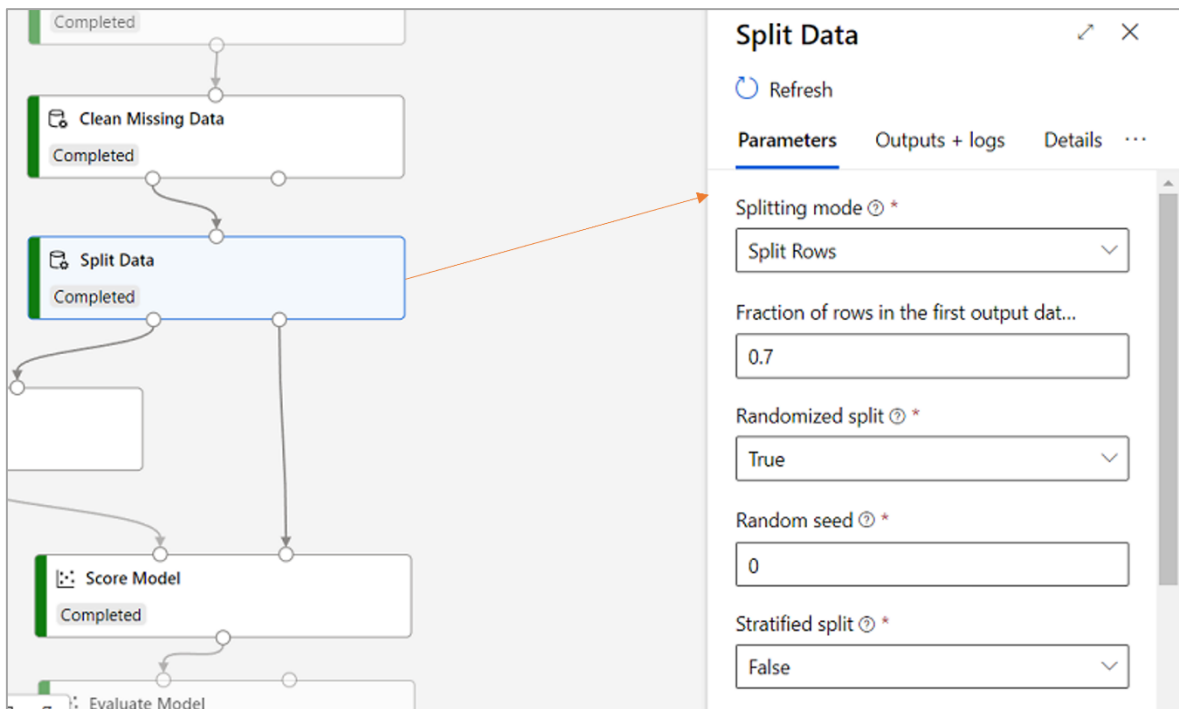


Kuva 22. Datan puhdistaminen

5.5 Mallin kouluttaminen

Mallin kouluttamiseen käytetään regressiota, koska halutaan ennustaa numeerisessa muodossa olevaa lämpötilaa. Tässä kokeilussa käytetään lineaarista regressiota (Linear Regression). Se soveltuu hyvin myös mahdollisesti jatkossa sovelluskohteista saadun datan kouluttamiseen. Mallin kouluttamisvaiheessa datan pilkkominen on hyvin yleinen tehtävä. Split Data -moduuli on raahattavissa suunnittelijassa vasemmalta myös Data Transformation -otsikon alta ja Linear Regression -algoritmi otsikon Machine Learning Algorithms alta. Vasen ulostuloportti Clean Missing Data -moduulista tulee yhdistää Split Data -moduuliin, koska oikea portti sisältää hylättyä dataa.

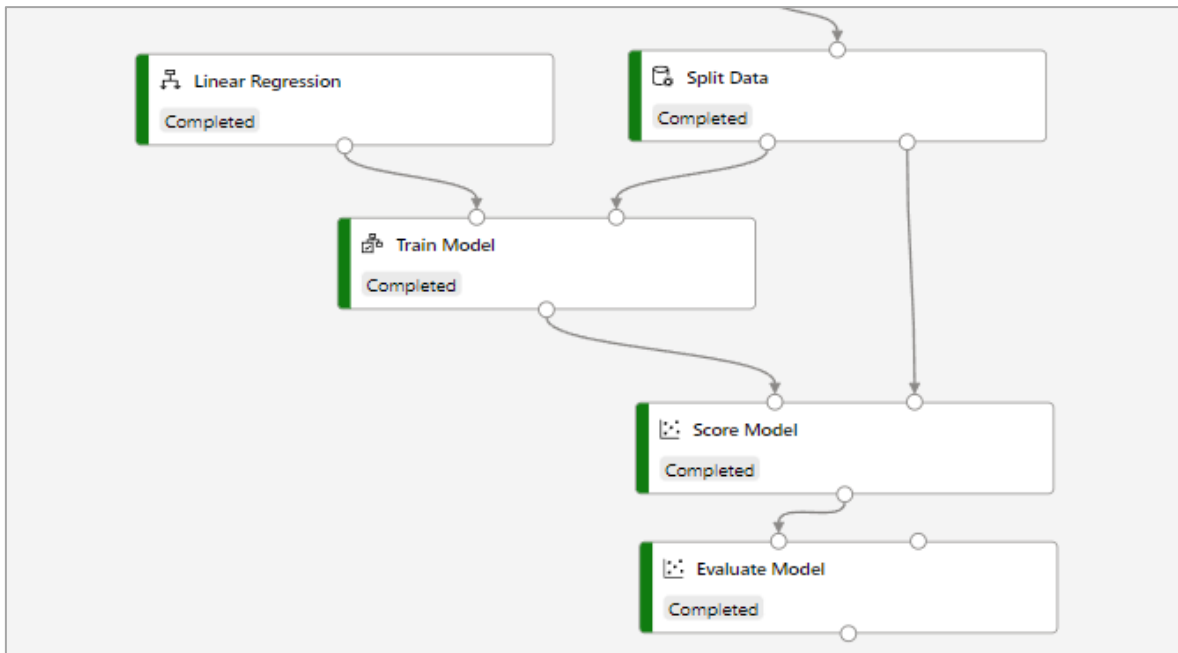
Asetetaan Split Data -moduulin tietoihin oikealle Fraction of rows in the first output dataset arvoksi 0.7 (Kuva 23). Se tarkoittaa, että data jaetaan satunnaisesti niin, että koulutusdatan osuus on 70 % ja testidatan 30 %. Raahataan tämän jälkeen Model Training -otsikon alta vasemmalta Train Model kanvaalle ja yhdistetään sen vasempaan sisääntuloporttiin Linear Regression -moduuli ja oikeaan Split Data -moduuliin sen vasemmasta ulostuloportista, koska se sisältää harjoitusdatan. Split Datan oikea ulostuloportti sisältää testidatan. Valitaan vielä Train Model -moduulin tietoruudusta oikealla Edit column, josta avautuu Label column -ikkuna. Valitaan alasvetovalikosta Column names ja kirjoitetaan tekstikenttään lämpötila täsmälleen siinä kirjoitusasussa kuin se esiintyy datasetissä, tässä kokeilussa DewPointCelsius ja tallennetaan lopuksi valinta.



Kuva 23. Datat pilkkominen osiin

5.6 Mallin arviointi

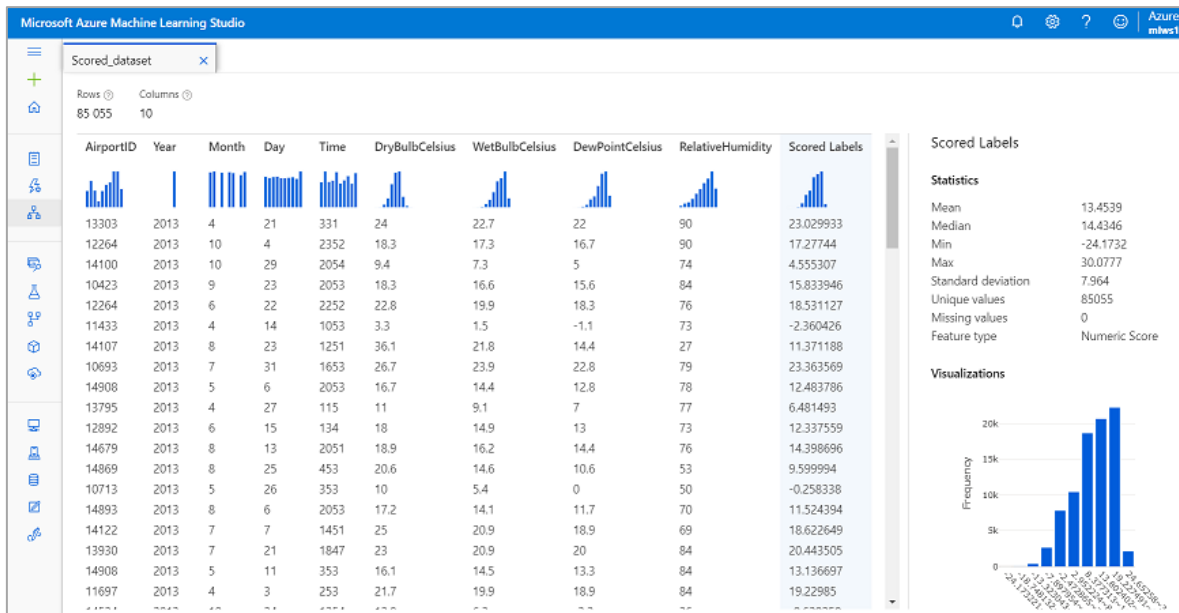
Kun malli on koulutettu käyttämällä 70 prosenttia datasta, voidaan sitä käyttää lopun 30 prosentin pisteytykseen ja mallin toimivuuden testaamiseen. Raahataan vasemmalta valikosta Model Scoring & Evaluation alta Score Model oikealle kanvaalle. Yhdistetään Train Model -moduulin ulostuloportti Score Model -moduulin vasempaan sisääntuloporttiin (Kuva 24). Liitetään Split Data -moduulin testidatan oikea ulostuloportti Score Model -moduulin oikeaan sisääntuloporttiin. Evaluate Model löytyy myös vasemmalta valikosta Model Scoring & Evaluation alta ja raahataan se oikealle kanvaalle. Liitetään Score Model yhteen Evaluate Model -moduulin vasempaan sisääntuloporttiin.



Kuva 24. Mallin kouluttaminen

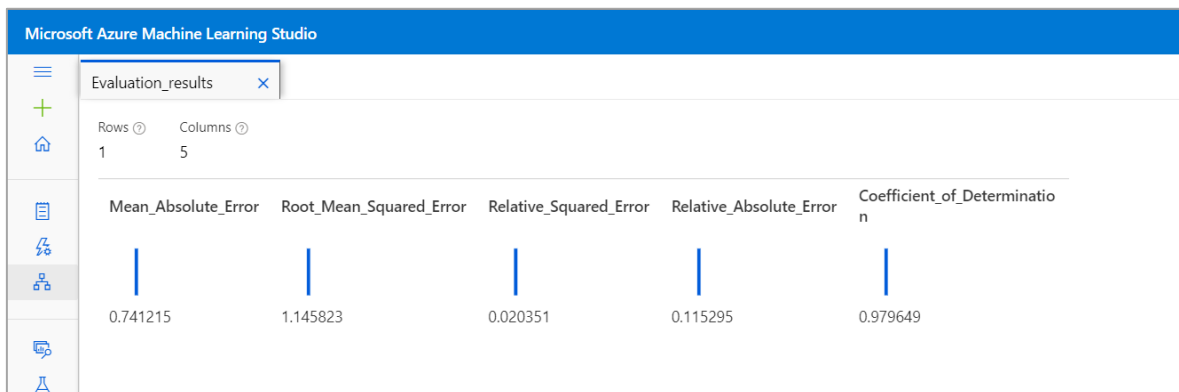
Kun kaikki osat on lisätty koneoppimisprosessiin, tehdään ajo mallin kouluttamiseksi. Valitaan kanvaan yläosasta Submit, joka näkyy kuvassa 17. Valitaan Set up pipeline run -valintaikkunasta Create new. Annetaan New experiment kohdassa Name nimi, joka tässä koikeilussa voisi olla esimerkiksi WeaaherTemperatures. Lopuksi valitaan Submit. Ajon tila ja yksityiskohdat näkyvät kanvaan oikeassa yläkulmassa. Ensimmäinen ajo voi kestää jopa 20 minuuttia, mutta myöhemmin ajot tapahtuvat nopeammin.

Onnistuneen ajon jälkeen voidaan tarkastella tuloksia. Painamalla Score Model -moduulin päällä hiiren kakkospainiketta ja valitsemalla Preview data > Scored dataset sekä avautuvasta pisteytetystä datasetistä Scored Labels nähdään tulokset (Kuva 25). Tilastotiedoista voidaan nähdä muun muassa lämpötilan keskiarvo, mediaani arvo eli keskimäinen arvo, pienin ja suurin arvo sekä lämpötilojen lukumäärä diagrammilla kuvattuna.



Kuva 25. Pisteytetty data

Mallia arvioimalla nähdään, kuinka hyvin koulutettu malli toimi testidatalla. Painamalla Evaluate Model -moduulin päällä hiiren kakkospainiketta ja valitsemalla Preview data > Evaluation results nähdään tulokset (Kuva 26).



Kuva 26. Mallin arviointi

Tuloksissa näytetään viisi tilastollista tunnuslukua:

1. Mean Absolute Error (MAE): keskimääräinen absoluuttinen virhe, absoluuttisten virheiden keskiarvo. Virhe on ennustetun arvon ja todellisen arvon välinen ero.
2. Root Mean Squared Error (RMSE): keskineliövirheen neliöjuuri, testiaineistolla tehtyjen ennusteiden neliövirheiden keskiarvon neliöjuuri.

3. Relative Absolute Error: suhteellinen absoluuttinen virhe, absoluuttisten virheiden keskiarvo suhteessa todellisten arvojen ja kaikkien todellisten arvojen keskiarvon absoluuttiseen eroon.
4. Relative Squared Error: suhteellinen neliövirhe, joka on neliövirheiden keskiarvo suhteessa todellisten arvojen ja kaikkien todellisten arvojen keskiarvon väliseen neliöeroon.
5. Coefficient of Determination: määrityskerroin, joka tunnetaan myös nimellä R-neliöarvo ja osoittaa, kuinka hyvin malli sopii dataan.

Mitä pienempi kunkin virhettä kuvaavan tunnusluvun arvo on, sitä parempi. Mitä pienempi arvo, sitä lähempänä ennusteet ovat todellisia arvoja. Mitä lähempänä määrityskerroin on yhtä (1.0), sitä paremmat ennusteet ovat. (Microsoft 2021j.)

Tässä työssä saatiin ensimmäisellä ajokerralla huomattavan suuret virhearvot ja määrityskerroin oli vain 0.33. Mallin parantamiseksi otettiin malliin lisää sarakkeita datasetistä kuten luvussa 5.4 aiemmin kerrottiin. Toisella ajokerralla malli saatiin koulutettua paremmin. Virhettä kuvaavat tunnusluvut saatiin pieniksi ja määrityskertoimeksi noin 0.98. Voidaan siis todeta, että ennusteet saatiin jo lähelle todellisia arvoja ja sikäli malli sopii hyvin dataan.

Seuraavana askeleena malli voidaan ottaa käyttöön reaaliaikaisella päätepisteellä (endpoint). Tämä tapahtuu muuttamalla koulutusputki reaaliaikaiseksi, jolloin poistetaan koulutusmoduuleja ja lisätään verkkopalvelun syötteitä ja ulostuloja pyyntöjen käsittelemiseksi (Microsoft 2021l). Lisäksi Azure Machine Learning -malleja voidaan käyttää Power BI:ssä tarkemmin raportointiin ja visualisointiin muodostamalla yhteys malliin Power Query Editorissa (Microsoft 2021m).

6 Yhteenveto ja pohdinta

Tämän opinnäytetyön tarkoituksena oli tutkia koneoppimista Azure Machine Learning -palvelussa ja selvittää palvelun hyödyntämistä tietojenkäsittelyn opetuksessa. Työssä pyrittiin vastaamaan kysymyksiin, mitä Azure Machine Learning on, mitä palvelun käyttö edellyttää ja miten palvelua voidaan hyödyntää opetuksessa. Tutkimuskysymyksiin saatiin vastaukset tutustumalla Azure Machine Learning -palveluun, aiempaan kirjallisuuteen ja tutkimuksiin. Työssä hyödynnettiin myös Microsoftin aineistoja ja oppaita. Opinnäytetyössä saavutettiin tavoitteet ja tutkimuskysymyksiin saatiin vastukset.

Teoriaosuudessa tarkasteltiin koneoppimista käsitteenä ja palvelua teknologian näkökulmasta. Toiminnallisen osuuden lopputuloksena syntyi ohjeistus vaihe vaiheelta Azure Machine Learning -ympäristön käyttöönotosta valmiin opetetun koneoppimismallin kehittämiseen, datan analysointiin ja visualisointiin. Työn lopputuloksena saatiin kokemusperäistä tietoa ja syvällisempää ymmärrystä palvelusta ja sen käytettävyydestä opetuksessa. Ohjeistusta voidaan hyödyntää opetuksessa koneoppimisen opettamisessa ja kokeilemisessa. Sen lisäksi, että opinnäytetyöstä on hyötyä toimeksiantajalle, se on hyödyllinen koneoppimisesta kiinnostuneille opiskelijoille sekä muille Azuren koneoppimisympäristöstä kiinnostuneille. Työ myös rohkaisee kokeilemaan koneoppimista, vaikkei olisi koodausosaamista. Työn tekijänä sain arvokasta tietoa ja kokemusta koneoppimisesta ja sen toteuttamisesta.

Koska lähdekirjallisuus on pääasiassa englanninkielistä, joidenkin suomenkielisten vastineiden löytäminen oli haastavaa. Alan nopea kehittyminen vaati myös tarkkuutta relevantin lähdemateriaalin löytämisessä. Vaikka tarkoituksena ei varsinaisesti ollut optimaalisen mallin kehittäminen, vaan palvelun kokeileminen oppimis- ja opetustarkoituksessa, ennusteet saatiin lähelle todellisia arvoja ja luotu malli sopii hyvin dataan.

Työssä käytettiin ohjelman tarjoamaa valmista datasettiä, joka osoittautui hyvin laajaksi, ja sen muokkausmahdollisuudet olivat rajalliset Azure Machine Learning Studiossa ja Designerilla. Koska virtuaalikoneaika on kallista koneen valintaan ja optimaalisen koneen löytämiseen kannattaa panostaa. Datasetin koko myös vaikuttaa käytettävään koneaikaan. Ensimmäiset kokeilut kannattaisikin tehdä pienemmällä datasetillä. Toisaalta mitä suurempi datamäärä, sitä helpompi on saada malli koulutettua vastaamaan dataa. Huomiota kannattaa siis käyttää datan esikäsittelyyn. Usein se onkin työläin ja eniten aikaa vaativa vaihe tekoäly- ja koneoppimiskehityksessä. Sitä kuvaa hyvin, että esimerkiksi Tesla on palkannut tuhat ihmistä ainoastaan nimeämistyöhön (Lähteenmäki 2021).

Oppimisen ja opetuksen näkökulmasta kooditon Designer-suunnittelutyökalu oli helppo ottaa käyttöön, ja sen avulla pääsi nopeasti alkuun koneoppimisessa. Käytännön

sovelluksissa Designerin soveltuvuutta kannattaa kuitenkin arvioida tapauskohtaisesti ja koodillinen mallin rakentaminen Notebooksilla tai automaattisella Automated ML:lla voivat tulla kysymykseen. Tässä työssä mallin kouluttamiseen käytettiin regressioalgoritmia, koska se on yleisesti käytetty numeerisen datan ennustamiseen ja siitä on hyötyä jatkokehityksessä esimerkiksi koneoppimisen suorittamisessa IoT-sovelluksissa.

Seuraavan kolmen vuoden aikana koneoppimisen tärkeimmät trendit ja kehityssuunnat ovat koneoppimisen yhdistäminen IoT:iin, automaattinen koneoppiminen, parempi kyberturvallisuus sekä koneoppimisen ja tekoälyn etiikka. Koneoppimisen ja IoT:n yhdistäminen on herättänyt paljon keskustelua ja on kauan odotettu trendi. Se liittyy 5G:n kehitykseen ja käyttöön, josta tulee alusta esineiden internetin kehittämiseksi. Koska 5G tuo mukanaan suuria nopeuksia, laitteet reagoivat nopeasti ja siirtävät ja vastaanottavat enemmän tietoa. IoT-teknologia mahdollistaa useiden laitteiden yhdistämisen yhteen verkkoon internetin avulla. (Komarraju 2021; Musienko 2021.)

Koneoppimisessa riittää myös jatkokehittämiskohteita. Koneoppimisen käytöllä on havaittu olevan hyötyä yrityksille muun muassa kannattavuuden kasvuna. Yhtenä kehityskohteena onkin koneoppimiskäytön kehittäminen yritysten tarpeisiin esimerkiksi opiskeluprojekteissa. Koneoppimisen hyödyntäminen esineiden- tai teollisten esineiden internetin sovelluksessa on myös luonnollinen jatkumo tälle työlle. Niitä käsiteltiin työn tietoperustaosuudessa, muttei tämän työn puitteissa ollut mahdollista toteuttaa. Monitoroinnista ja ennustavasta huollosta saatavasta sensoridatan käsittely koneoppimisen menetelmillä olisikin hyvä jatkokehityskohde, samoin muiden koneoppimistekniikoiden kokeileminen ja koneoppimismallin kehittäminen koodin avulla. Opinnäytetyön muodossa voidaan myös tutustua, miten erilaista dataa tuodaan palveluun eri tietolähteistä analysoidavaksi ja visualisoida Power BI:ssä. Koneoppimiseen voitaisiin mennä myös syvemmälle kuten käsitellä kuvia syväoppimisen menetelmillä.

Lähteet

- Airaksinen, T. 2009. Toiminnallisen opinnäytetyön kirjoittaminen. Viitattu 23.6.2021. Saatavissa <https://www.slideshare.net/TiinaMarjatta/toiminnallinen-opinnytety-tekstin>
- Baker, V., Elliot, B., Sicular, S., Mullen, A. & Brethenoux, E. 2021. Magic Quadrant for Cloud AI Developer Services. Gartner. Viitattu 11.8.2021. Saatavissa <https://www.gartner.com/doc/reprints?id=1-25C36W9W&ct=210226&st=sb>
- Barnes, J. 2015. Azure Machine Learning - Microsoft Azure Essentials. Washington: Microsoft Press.
- Brynjolfsson, E. & McAfee, A. 2017. The Business of Artificial Intelligence, What it can – and cannot – do for your organization. Harvard Business Review. The Big Idea Series / Artificial Intelligence, For Real 7/2017, 4–11.
- Collin, J. & Saarelainen A. 2016. Teollinen internet. Helsinki: Talentum.
- eCloudvalley. 2021. Introduction to Machine Learning: Is AutoML Replacing Data Scientists? Viitattu 14.9.2021. Saatavissa <https://www.ecloudvalley.com/mlintroduction/>
- Gollapudi, S. & Laxmikanth, V. 2016. Practical Machine Learning. Birmingham: Packt Publishing Ltd.
- Gupta, B. 2019. Three Things to Know About Reinforcement Learning. Viitattu 14.9.2021. Saatavissa <https://medium.com/@bhaskergupta/three-things-to-know-about-reinforcement-learning-2760700f0723>
- Harkut D. 2019. Cloud Computing - Technology and Practices. IntechOpen. Viitattu 20.6.2021. Saatavissa <https://www.intechopen.com/books/cloud-computing-technology-andpractices/introductory-chapter-cloud-computing>
- Helsingin yliopisto & Reaktor Innovations. 2021. Elements of AI. Verkkokurssi. Viitattu 19.8.2021. Saatavissa <https://course.elementsofai.com/fi>
- Hyppönen, M. 2018. Pilven reunalla. Fujitsun blogi 19.2.2018. Viitattu 2.9.2021. Saatavissa <https://www.blogi.fujitsu.fi/2018/02/pilven-reunalla/>
- IteWiki. 2021. Koneoppiminen. Viitattu 22.7.2021. Saatavissa <https://www.itewiki.fi/opas/koneoppiminen/>
- Kapitanova, K. & Son, H. 2012. Machine Learning Basics. Teoksessa Hu, F. & Hao, Q. (toim.) Intelligent Sensor Networks: The Integration of Sensor Networks, Signal

Processing and Machine Learning. Boca Raton, the United States of America: CRC Press, 3–29.

Kaspersky Lab. 2020. Machine Learning Methods for Malware Detection. White Paper. Viitattu 17.6.2021. Saatavissa <https://media.kaspersky.com/en/enterprise-security/Kaspersky-Lab-Whitepaper-Machine-Learning.pdf>

Kavis, M. 2014. Architecting the cloud: Desing decisions for cloud computing service models (SaaS, PaaS and IaaS). Somerset: Wiley.

King, T. 2021. What's Changed: 2021 Gartner Magic Quadrant for Data Science and Machine Learning Platforms. Business Intelligence Solutions Review 4.3.2021. Viitattu 7.7.2021. Saatavissa <https://solutionsreview.com/business-intelligence/whats-changed-2021-gartner-magic-quadrant-for-data-science-and-machine-learning-platforms/>

Kokkarainen, I & Ala-Mutka, K. 2002. Tietorakenteet ja Algoritmit. 2. painos. Helsinki: Satku.

Komarraju, A. 2021. Top Machine Learning Trends for 2022 and Beyond. Viitattu 27.10.2021. Saatavissa <https://www.analyticsinsight.net/top-machine-learning-trends-for-2022-and-beyond/>

Kuznetsov, V. 2018. Machine Learning as a Service for HEP. ArXiv:1811.04492, HepEx, Physics: Physics. Viitattu 20.6.2021. Saatavissa <http://arxiv.org/abs/1811.04492>

LAB-ammattikorkeakoulu. Tietoa meistä. Viitattu 23.6.2021. Saatavissa <https://lab.fi/fi/info/tietoa-meista>

Lee, W-M. 2019. Python Machine Learning. Indianapolis, Indiana: Wiley.

Lähteenmäki, P. 2021. Tulevaisuus on Teslan. Talouselämä 33/2021, 22–29.

Meldrum, N. 2017. Factories of the future: Meeting the challenge of Industry 4.0. Microsoft Industry Blogs. Viitattu 26.8.2021. Saatavissa <https://cloudblogs.microsoft.com/industry-blog/manufacturing/2017/12/12/the-factory-of-the-future-meeting-the-challenge-of-industry-4-0/>

Merilehto, A. 2018. Tekoäly - Matkaopas johtajalle. Helsinki: Alma Talent Oy.

Microsoft. 2019. Tutorial: An end-to-end solution using Azure Machine Learning and IoT Edge. Viitattu 2.9.2021. Saatavissa <https://docs.microsoft.com/en-us/azure/iot-edge/tutorial-machine-learning-edge-01-intro?view=iotedge-2020-11>

Microsoft. 2020. How to select algorithms for Azure Machine Learning. Viitattu 7.9.2021. Saatavissa <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-select-algorithms>

Microsoft. 2021a. What is machine learning? Viitattu 16.6.2021. Saatavissa <https://azure.microsoft.com/en-us/overview/what-is-machine-learning-platform/>

Microsoft. 2021b. What is Azure Machine Learning? Documentation. Viitattu 21.6.2021. Saatavissa <https://docs.microsoft.com/en-us/azure/machine-learning/overview-what-is-azure-ml>

Microsoft. 2021c. Machine learning algorithms - An introduction to the math and logic behind machine learning. Viitattu 20.7.2021. Saatavissa <https://azure.microsoft.com/en-us/overview/machine-learning-algorithms/#overview>

Microsoft. 2021d. Machine Learning Algorithm Cheat Sheet for Azure Machine Learning designer. Viitattu 17.8.2021. Saatavissa <https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-cheat-sheet>

Microsoft. 2021e. Build in the cloud with an Azure free account. Viitattu 16.9.2021. Saatavissa https://azure.microsoft.com/en-gb/free/?WT.mc_id=A261C142F

Microsoft. 2021f. Quickstart: Create workspace resources you need to get started with Azure Machine Learning. Viitattu 16.9.2021. Saatavissa <https://docs.microsoft.com/en-us/azure/machine-learning/quickstart-create-resources>

Microsoft. 2021g. Azure Machine Learning pricing. Viitattu 19.9.2021. Saatavissa <https://azure.microsoft.com/en-gb/pricing/details/machine-learning/#pricing>

Microsoft. 2021h. What is Azure? Viitattu 7.10.2021. Saatavissa <https://docs.microsoft.com/en-us/learn/modules/intro-to-azure-fundamentals/what-is-microsoft-azure>

Microsoft. 2021i. Example pipelines & datasets for Azure Machine Learning designer. Viitattu 15.10.2021. Saatavissa <https://docs.microsoft.com/fi-fi/azure/machine-learning/samples-designer>

Microsoft. 2021j. Tutorial: Designer - train a no-code regression model. Viitattu 17.10.2021. Saatavissa <https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-designer-automobile-price-train-score>

- Microsoft. 2021k. Connect to data with the Azure Machine Learning studio. Viitattu 24.10.2021. Saatavissa <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-connect-data-ui#create-datasets>
- Microsoft. 2021l. Tutorial: Designer - deploy a machine learning model. Viitattu 26.10.2021. Saatavissa <https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-designer-automobile-price-deploy>
- Microsoft. 2021m. Tutorial: Consume Azure Machine Learning models in Power BI. Viitattu 26.10.2021. Saatavissa <https://docs.microsoft.com/en-us/power-bi/connect-data/service-aml-integrate?context=azure/machine-learning/context/ml-context>
- Musienko, Y. 2021. Top Machine Learning Trends and Technologies for 2022. Viitattu 27.10.2021. Saatavissa <https://merehead.com/blog/top-machine-learning-trends-technologies-2022/>
- Pardeshi, S. 2021. Increase ROI and business impact with Azure Machine Learning. Viitattu 26.10.2021. Saatavissa <https://azure.microsoft.com/en-gb/blog/increase-roi-and-business-impact-with-azure-machine-learning/>
- Ranger, S. 2018. What is cloud computing? Everything you need to know about the cloud explained. Viitattu 20.6.2021. Saatavissa <https://www.zdnet.com/article/what-is-cloud-computing-everything-you-need-to-know-about-the-cloud/>
- Ruokolainen, L. 2018. Ylisovittaminen ja kuinka sen kanssa voi tulla toimeen. Bilot Group Blogi 11.04.2018. Viitattu 20.11.2021. Saatavissa <https://bilot.group/articles/yliovittaminen-ja-kuinka-sen-kanssa-voi-tulla-toimeen/>
- Samuel, A. L. Some studies in machine learning using the game of checkers. IBM Journal of Research and Development. Vol. 44, Iss. 1/2 (Jan/Mar 2000), 206–226.
- Tutorialspoint. 2021. Types of Learning. Viitattu 21.7.2021. Saatavissa https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_types_of_learning.htm
- Vilkka, H. & Airaksinen, T. 2003. Toiminnallinen opinnäytetyö. Helsinki: Tammi.
- Wolhuter, S. 2021. Machine Learning as a Service — What is it? Who are the big players? WeAreBrain Blog 26.3.2021. Viitattu 20.6.2021. Saatavissa <https://wearebrain.com/blog/ai-data-science/machine-learning-as-a-service-mlaas/>

Liite 1. Machine Learning Algorithm Cheat Sheet

