

# **Data-analysointi ja tekoälyalgoritmikehitys automaatioalalla**

Johtava tutkija, Mika-Petri Laakkonen

21.5.2025

# Tekoälyalgoritmimallintamisen paradigmat

1. Neuroverkkojen mallintaminen: Perustuu ihmisen aivoihin ja muistiin, käyttäen historiallisia tietoja kuvioden tunnistamiseen ja päätöksentekoon.

2. Geneettinen mallintaminen: Perustuu biologiaan ja genetiikkaan ja keskittyy evoluutioprosesseihin monimutkaisten ongelmien ratkaisemiseksi.

3. Genealoginen mallintaminen: Populaatiodynamiikkaan perustuvien pitkän aikavälin mukautumisprosessien huomioiminen.

4. Fysikaalisten ilmiöiden mallintaminen: Todellisen maailman fysiikan inspiroima, jossa tekoäly nähdään dynaamisena, itseorganisoituvana järjestelmänä.

# Koneoppiminen neuroverkoilla/Fysiikan Nobel 2024

- John Hopfieldin 1982 kehittämä neuroverkkomallin (ANN) “Hopfieldin neuroverkot” palkittiin fysiikan Nobelilla. Tosin nuo Hopfieldin neuroverkkomallit pohjautuvat enemmän juuri tuohon ensimmäiseen paradigmaan (neuroverkkojen mallintaminen) eikä varsinaisesti fysiikkaan.
- Fysiikkaan pohjautuvat mallit näkyvät vasta 50 vuoden aikajänteellä ja ehkä sen jälkeen fysiikan Nobel pohjautuu neljänteen paradigmaan eli fysikaalisten ilmiöiden mallintamiseen
- Ensimmäiset fysiikkaan perustuvat mallit tulivat tekoälytutkimukseen vasta 2018. Tuolloin toimin vierailevana professorina Sorbonnen yliopistossa (2018 -2019) , mutta nekin mallit, mitä kehitimme pohjautuivat suurimmaksi osaksi neuroverkkoihin, vaikka tutkittavana ilmiönä olivatkin fysiikkaan pohjautuvat ilmiöt mm. turbulenssi.



John J. Hopfield  
Nobel Prize in Physics 2024

Born: 15 July 1933, Chicago, IL, USA

Affiliation at the time of the award: Princeton University,  
Princeton, NJ, USA

Prize motivation: “for foundational discoveries and inventions that enable machine learning with artificial neural networks”

# Koneoppiminen neuroverkoilla

- John Hopfieldin vuonna 1982 esittelemä Hopfield-verkko on eräänlainen toistuva neuroverkko, joka on suunniteltu toimimaan sisällöllisesti ohjattavana muistijärjestelmänä – eli se voi hakea tallennettuja kuvia osittaisen tai kohinaisen syötteen perusteella.
- Toistuvat ja täysin kytketyt neuroverkot
- Binääriset kynnyssolmut
- Energian minimointi
- Ei itsesilmukoita
- Hopfield-verkko on neurolaskennan perustavanlaatuinen malli, joka havainnollistaa, kuinka muisti ja kuvien tunnistaminen voivat syntyä yksinkertaisista, biologisesti inspiroituneista periaatteista. Se loi pohjan myöhemmälle kehitykselle; assosiatiiviselle muistille ja neurodynamiikalle.



# Automaatioala

---

Automaatio tarkoittaa itsetoimivaa laitetta tai järjestelmää.

---

Automaation hyötyjä ovat: toistettavuus, tiukempi laadunhallinta, jätteiden vähentäminen, integraatio yritysten muiden järjestelmien kanssa, kasvanut tuotanto ja pienentynyt työvoiman tarve.

---

Usein käytetään automaatiota varten tehtyjä tietokoneita ja ohjelmoitavia logiikoita (Programmable Logic Control) ohjaamaan järjestelmän toimilaitteita sensoreilta saatujen tietojen perusteella.

---

Toinen automaatio muoto on testausautomaatio, jossa tietokone ohjelmoidaan matkimaan ihmisen toimintaa tietokoneohjelmistojen manuaalisen testauksen aikana. (ns. testausautomaatio-ohjelmistot), jonka tuottaman ohjelman avulla testaajan työ voidaan toistaa täsmälleen samalla tavalla uudelleen myöhemmin.

# Data-analyysin käyttö automaatioalalla

Teollisen prosessin data: suuri määrä dataa kerätään, tallennetaan ja prosessoidaan tarkoituksena automatisoida ja kontrolloida teollista tuotantoa, jakelua ja varastointia.

Päätarkoitus data-analyysille on optimoida prosesseja ja parantaa yrityksen kilpailukykyä.

Yrityksissä dataa analysoidaan liiketoiminnan paremman ymmärryksen ja ohjauksen toivossa. Analysoitavia alueita ovat asiakkaat, myynti, markkinointi, hinnoittelu ja riskinotto.

Teksti- ja rakenneddata ovat tärkeitä informaation lähteitä. Teksti- ja rakennedatan analyysi auttaa informaation suodattamisessa, etsimisessä ja louhimisessa.

Kuvadatan avulla voidaan etsiä, tunnistaa esineitä, analysoida ja luokitella paikkoja ja yhdistää kuvadataa muihin informaation lähteisiin.

# Data-analyysi



Data-analyysi muodostaa datasta korkeamman tason informaatioita ja malleja. Eli representaatioita hyödyllisten johtopäätösten tekemiseen.



Klassinen data-analyysi varmistaa vastaako ilmiöstä kerätty tieto ilmiöstä oletettua mallia.



Eksploratiivinen data-analyysi pohjautuu informaatioteoreettisiin menetelmiin (mm. informaation mittaamista ja siirtämistä, tiedonpakkaamista)



Bayesilainen data-analyysi on laskennallinen menetelmä, joka perustuu ilmiötä kuvaavan asiantuntijatiedon ja sen epävarmuuden sisällyttämisen analyysin tuloksena syntyvän mallin rakenteeseen (priorijakaumat)

# Data-analyysi

- Kvalitatiivinen data-analyysi perustuu ilmiöstä käsittelevään laadulliseen tietoon, jota ei ole saatu numeeriseen asuun.
- Tiedonlouhinta (data mining) pyrkii löytämään ilmiöön liittyvästä datasta kiinnostavia säännönmukaisuuksia

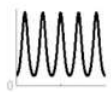


# Automaatioalan data-analyysi

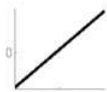
Five base kernels



Squared  
exp. (SE)



Periodic  
(PER)



Linear  
(LIN)

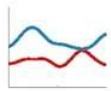


Constant  
(C)

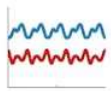


White  
noise (WN)

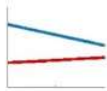
Encoding for the following types of functions



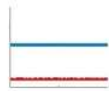
Smooth  
functions



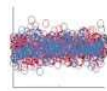
Periodic  
functions



Linear  
functions



Constant  
functions



Gaussian  
noise

- Tietokone käytti aikaisemmin enemmän aikaa operaation suorittamiseen kuin muistihakuun.
- Nykyään suorituksen kustannus on merkittävästi pienempi kuin tiedon muistista hakemisen kustannus.
- Ollakseen hyödyllinen algoritmin on ratkaistava yleinen täsmällisesti määritelty ongelma.
- Hyvällä algoritmilla on kolme ominaisuutta: oikeellisuus, tehokkuus ja toteutuksen helppous.
- Haastavimmat ongelmat algoritmeissa liittyvät optimointiin, jossa etsitään ratkaisua, joka maksimoi tai minimoi funktion.

# Automaatioalan data-analyysin vaiheet

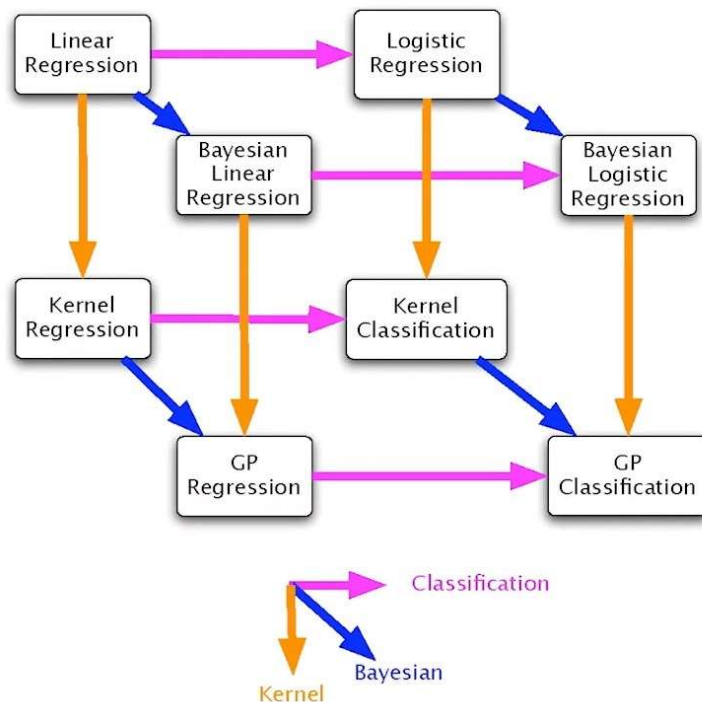
Valmistelu: suunnittelu, datan kerääminen, datan valinta

Esikäsittely: puhdistus, suodatus, täydentäminen, korjaus, standardointi, muuntaminen

Analysointi: visualisointi, korrelaatiot, yleistetty lineaarinen malli (regressio), ennustaminen, luokittelu, klusterointi

Jälkikäsittely: tulkinta, dokumentointi ja arviointi

# Gaussin, Bayesian ja Kernelin data-analyysi

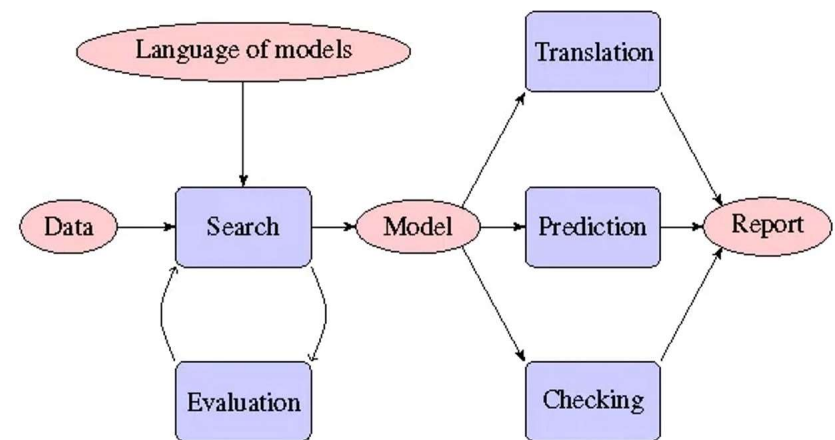


- Gaussin prosessiregressio: Funktioarvojen välisen kovarianssin määrittämiseen. (\*kovarienssi on tilastotieteessä kahden satunnaismuuttujan välisen riippuvuuden mitta)
- Bayesilainen optimointi on koneoppimismenetelmä: Tavoitefunktioiden mallintamiseen (mm. menetelmä on tarkoitettu suuntamaan näytteiden keruu etsintäavaruuden lupaaville alueille)
- Ydinmenetelmät: SVM:issä ja muissa algoritmeissa, jotka perustuvat samankaltaisuusmittoihin. Tätä menetelmää käytetään mm. kuvien analyysiin, klusterointiin, luokitteluun ja pääkomponenttianalyysiin)

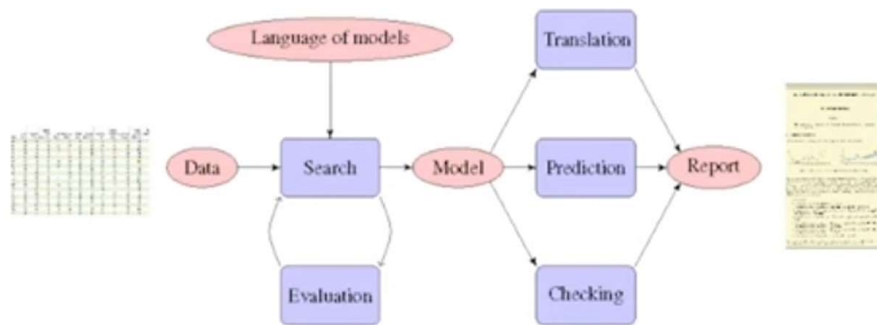
(\*regressioanalyysin avulla tutkitaan yhden tai useamman selittävän muuttujan yhteyttä selitettävään muuttujaan)

# Automaattinen tilastotieteilijä

- Dataa on nykyään kaikkialla, ja sen ymmärtämisestä, mallien rakentamisesta ja ennusteiden tekemisestä on suurta hyötyä automaatioalalla.
- Ongelma: data-analyytikoita, tilastotieteilijöitä ja koneoppimisen asiantuntijoita ei kuitenkaan ole tarpeeksi.
- Ratkaisu: Kehittää järjestelmä, joka automatisoi mallien löytämisen datasta, algoritmimalli(e)n valinnan sekä löydösten selittämisen käyttäjälle.



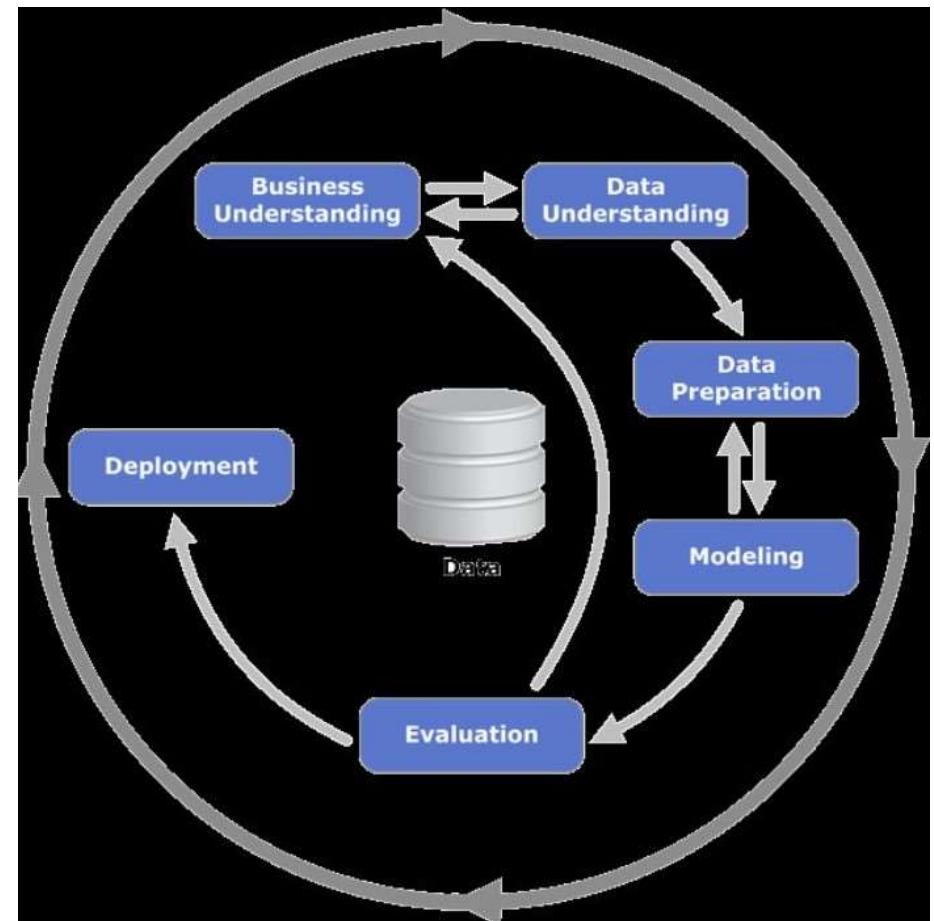
# Automaattisen tilastotieteilijän ainesosat



1. Algoritmimalli: riittävän ilmaisuvoimainen tallentaakseen reaailmaailman ilmiöitä ja ihmistilastotieteilijöiden käyttämiä tekniikoita
2. Hakumenetelmä: algoritmimallien tehokkaaseen tutkimiseen
3. Algoritmimallien arvioinnin periaatemenetelmä: algoritmimallin luotettavuus ja toistettavuus dataan analysointiin
4. Algoritmimallien automaattinen selittäminen: algoritmimalli oletusten tekeminen eksplisiittiseksi (selkeällä) tavalla, joka on ymmärrettävä myös muille kuin asiantuntijoille

# Koneoppimisen automatisointi

- Tiedonkeruun ja kokeiden suunnittelun automatisointi
- Ominaisuuksien valinnan, muuntamisen ja ymmärtämisen automatisointi
- Mallien löytämisen ja selittämisen automatisointi
- Laskentaresurssien allokoinnin automatisointi
- Päättelyn automatisointi





Dr. Mika-Petri Laakkonen  
Lead/Principal Research Scientist, Oulu University of Applied  
Sciences  
Full Professor, University of Haifa (Affiliated)  
Oulu University of Applied Sciences, Linnanmaa Campus  
Osoite/Address: Yliopistokatu 9, 90570 Oulu, FINLAND  
Mobile: +358 50 4688016  
Email: [mika-petri.laakkonen@oamk.fi](mailto:mika-petri.laakkonen@oamk.fi)

