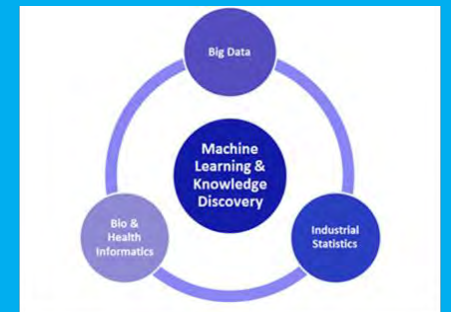


**UNIVERSITY
OF OULU**

Johdatus data-analytiikan, koneoppimisen ja tekoälyn hyödyntämiseen

Satu Tamminen
Oulun yliopisto / BISG
6.5.2019



Data Analysis and Inference Group



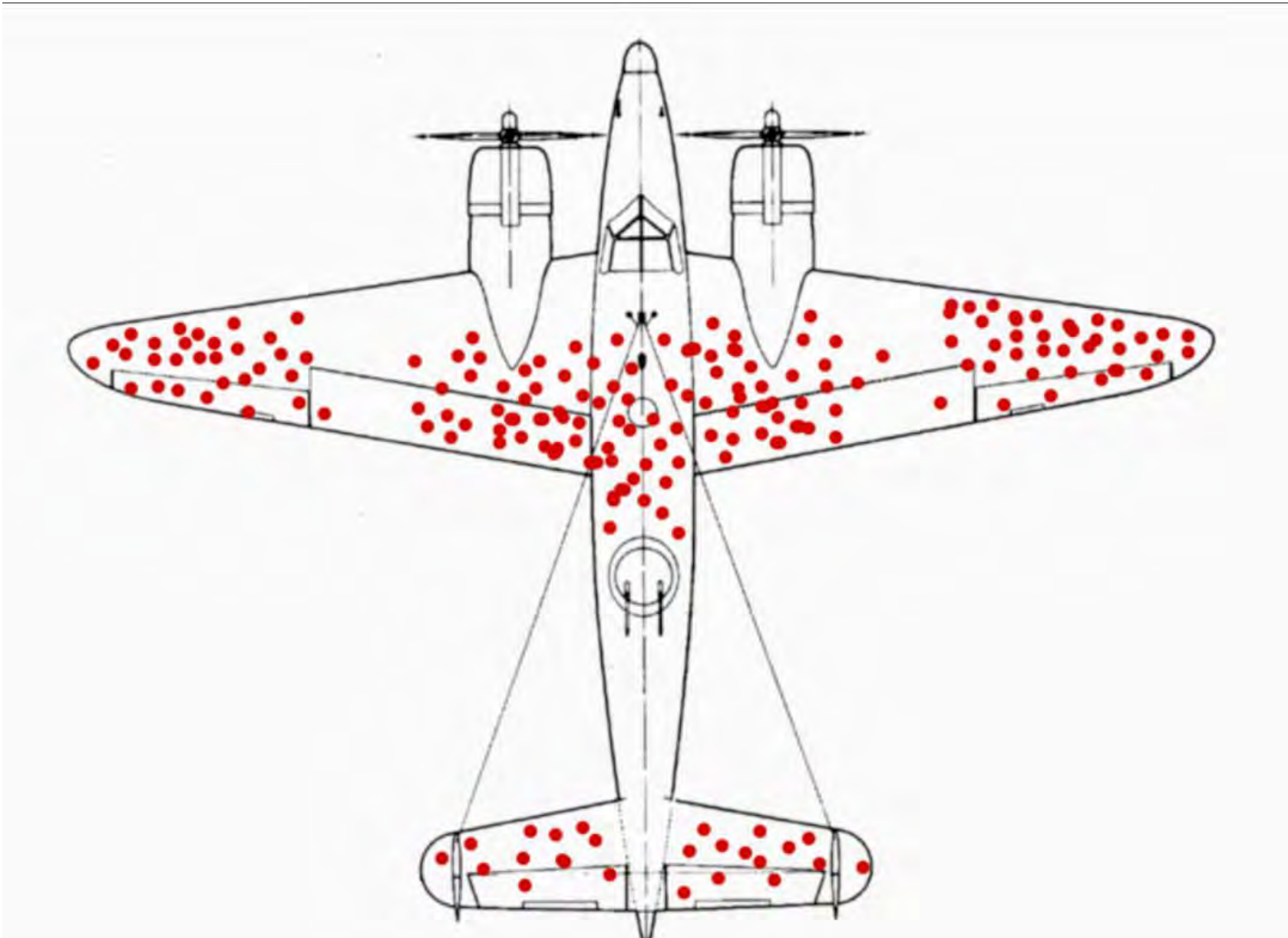
DATAN ROOLI KONEOPPIMISESSA JA TEKOÄLYSSÄ



Vale, emävale ja tilasto?



- Data ei valehtele, mutta tarinan voi kertoa väärin
- **Malli oppii datan perusteella:**
 - Sisältääkö data tietoa tutkittavasta asiasta?
 - Onko kyseessä edustava otos tutkitusta populaatiosta?
 - Onko mittausväli ajallisesti riittävän pitkä tai muuttujien mahdollisen vaihteluvälin kattava?
 - Sisältyykö mittauksiin epävarmuustekijöitä?





TEKOÄLY JA KONEOPPIMINEN



Käsitteitä

TEKOÄLY

- Älykkyyys tuottaa järkevää toimintaa
- Älykäs kone, laite, ohjelma, järjestelmä tai palvelu osaa valita parhaan mahdollisen toimintaratkaisun tilanteen mukaan.

Russel and Norvig: Artificial Intelligence: A Modern Approach

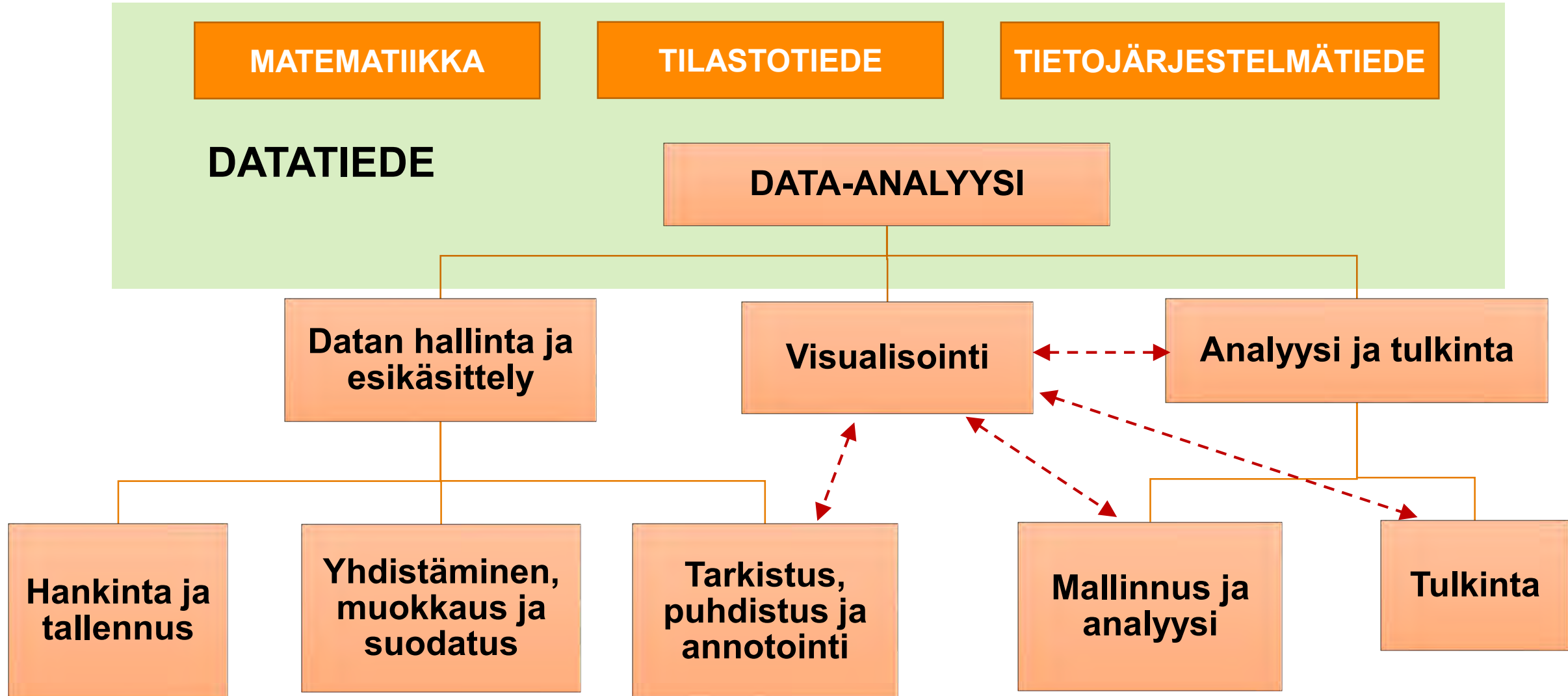
KONEOPPIMINEN

- Tietokonetekniikan osa-alue, jossa yleensä käytetään tilastotieteen menetelmiä
- Antaa tietokoneille kyvyn ”oppia” **datasta** (parantaa suorituskyykyään tietyn tehtävän suorittamisessa) ilman eksplisiittistä ohjelmointia.

VTT: Tekoälyn käsitekartta

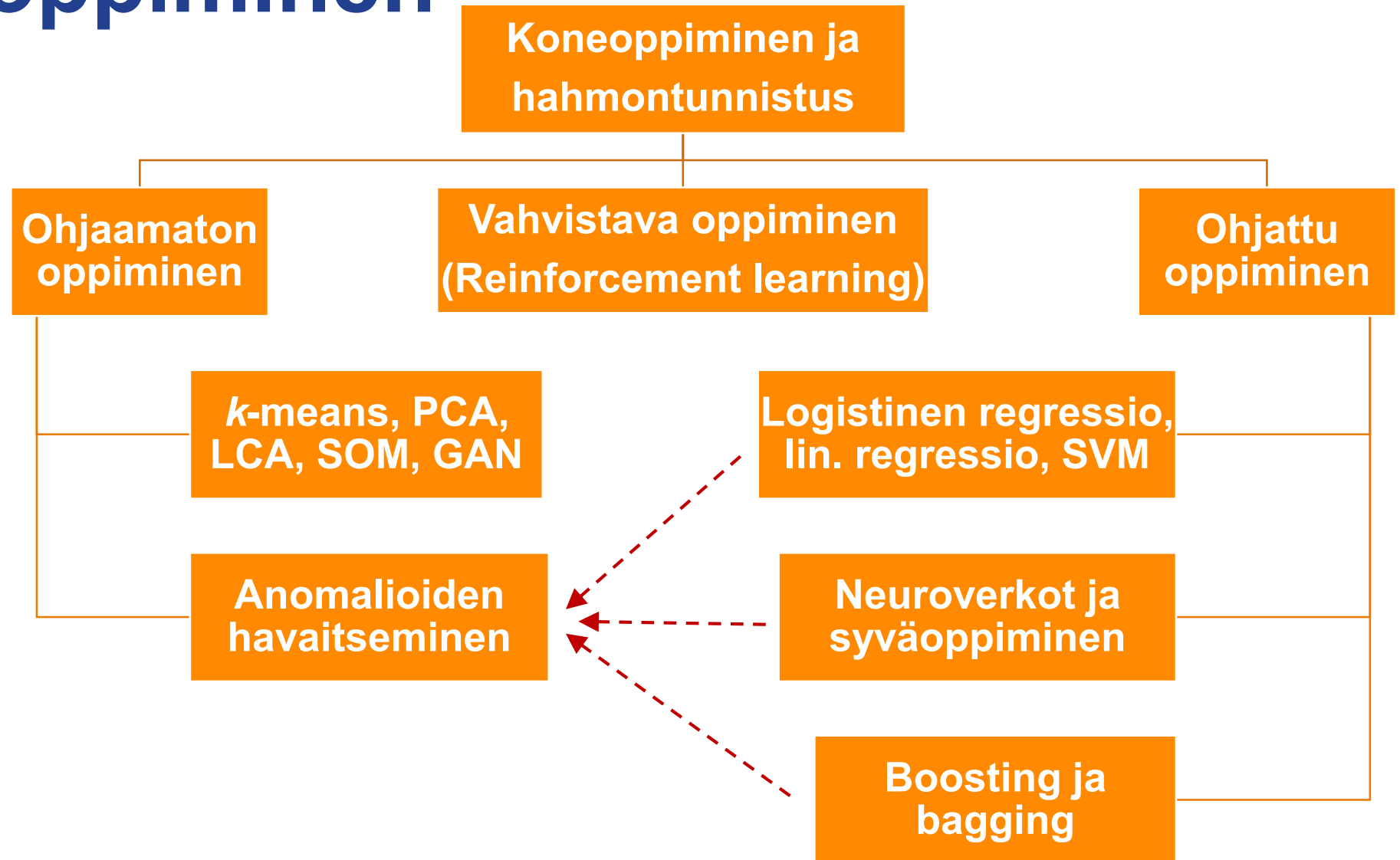


Data-analyysi





Koneoppiminen





DATAN ESIKÄSITTELY



Miksi?



Kun dataa kerätään eri lähteistä se voi olla

– **Puutteellista**

- Mittausarvoja puuttuu
- Kiinnostavia muuttujia puuttuu
- Raakamittausten sijaan on tallennettu vain aggregoituja arvoja

– **Kohinaista**

- Virheellisiä arvoja
- Outliereita

– **Yhteensopimatonta**

- Eri formaateissa
- Samoja muuttujia eri nimillä

– **Biasoitunutta, epänormaalia tai täysin irrelevanttia**

– **Esikäsittely parantaa datan laatua**



Datan esikäsittely

- **Datan määrä ei korvaa laatua**
- **Garbage in = garbage out**
 - Datan analysointi onnistuu, kun data on hyvin esikäsitelty
- **More is more**
- **Kuinka paljon tulisi käyttää aikaa datan esikäsittelyyn?**
 - Mitä siitä seuraa, jos käyttää huonoa dataa?

ESIKÄSITTELYN TEHTÄVIÄ:

- Eri tietolähteiden integrointi yhteen
- Datan siivous
- Annotointi
- Transformointi
- Datamäärän tiivistäminen





Esikäsittely opettaa



Kun data valmistellaan hyvin analysointia tai malleja varten, myös datan käyttäjä oppii!

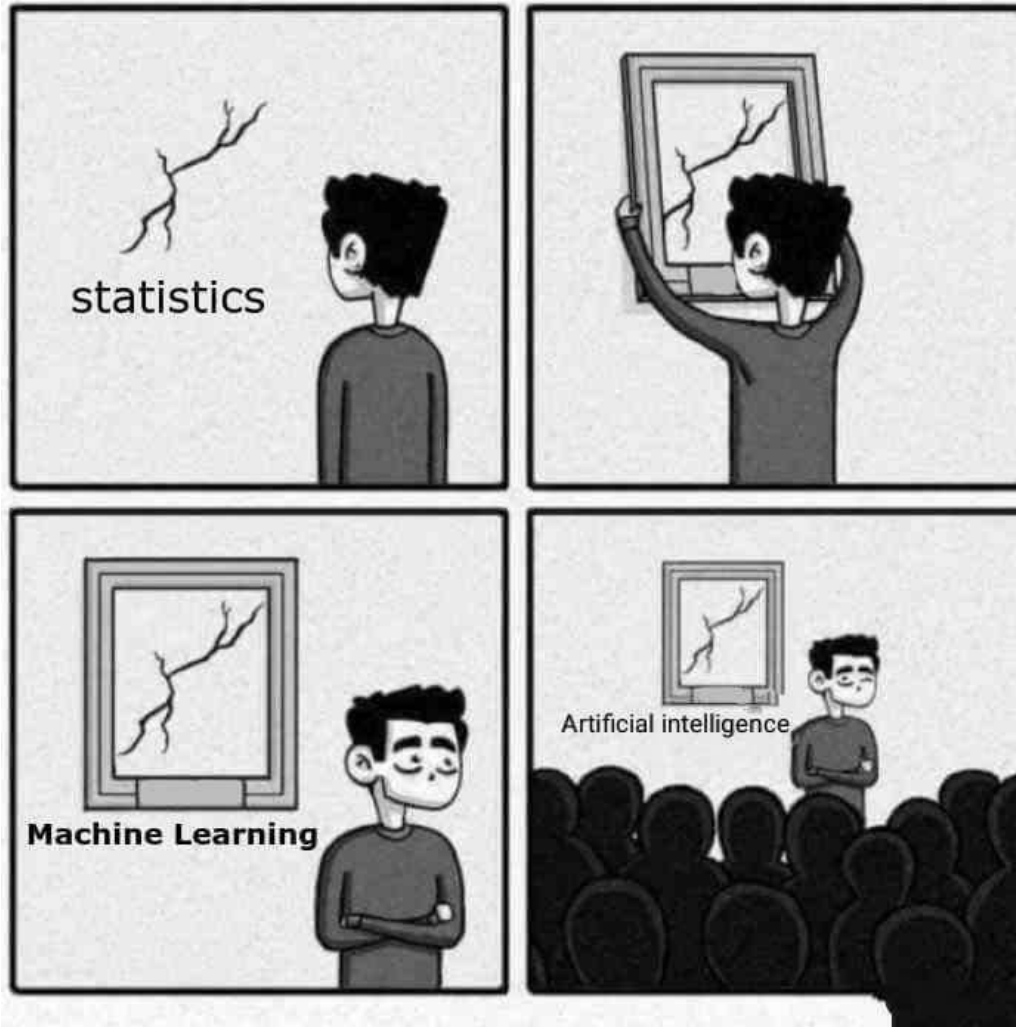
- **Datan esikäsittely varmistaa, että analysointimenetelmät ja mallit toimivat oikein**
- **Käyttäjä oppii tuntemaan sovellusalueen ja varmistuu, että on valittu oikeat menetelmät ja mallit**



TEKOÄLY PÄÄTÖKSENTEON TUKENA



Koneoppimisen hyödyt



- **Koneoppimisen huimat mahdollisuudet**
 - Laskentavoiman kehittyminen
 - Uudet ja entistä monimutkaisemmat algoritmit
 - Lähes loppumaton datan määrä
- **Monimutkaisten, epälineaaristen järjestelmien mallinnus on perinteisillä tilastollisilla menetelmillä vaikeaa tai mahdotonta**
- **Muuttuvat ympäristöt vaativat vuorovaikutusta, dynaamisuutta ja jatkuvaa oppimista**



Tulosten tulkinta



- Mallin hyvyys ei ole riittävä mittari kuvaamaan tulosten luotettavuutta
- Keskimääräinen hyvyys ei kerro yksittäisen ennusteen hyvyydestä
- Jos tiedät, mihin ennuste perustuu, opit ilmiöstäkin enemmän
- Voit paremmin tunnistaa opetusdataan liittyviä ongelmia, ja korjata ne
- Black box –malleista läpinäkyvyyteen



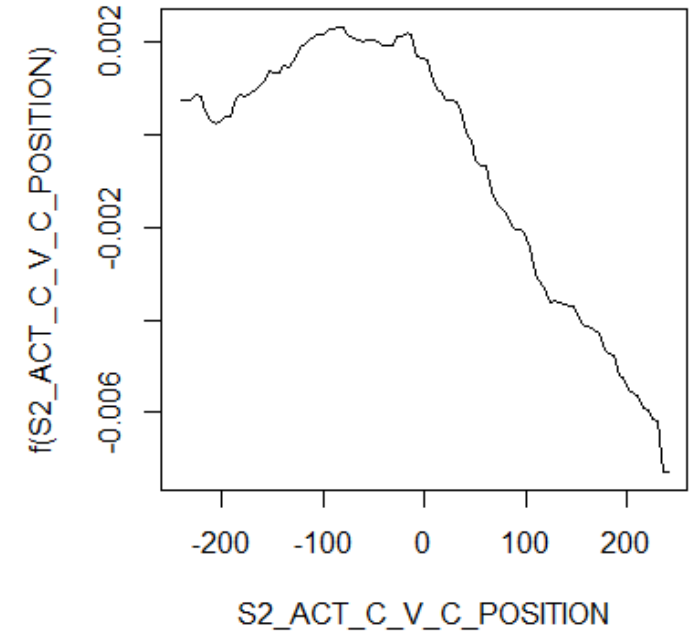
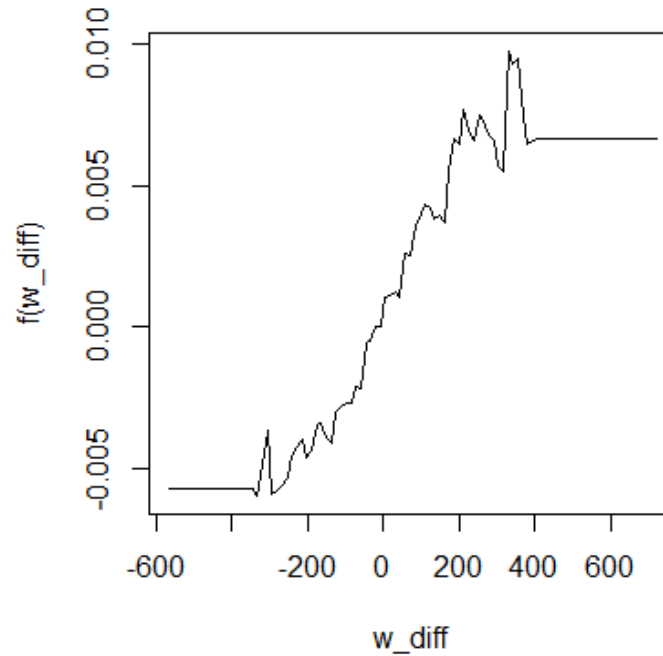
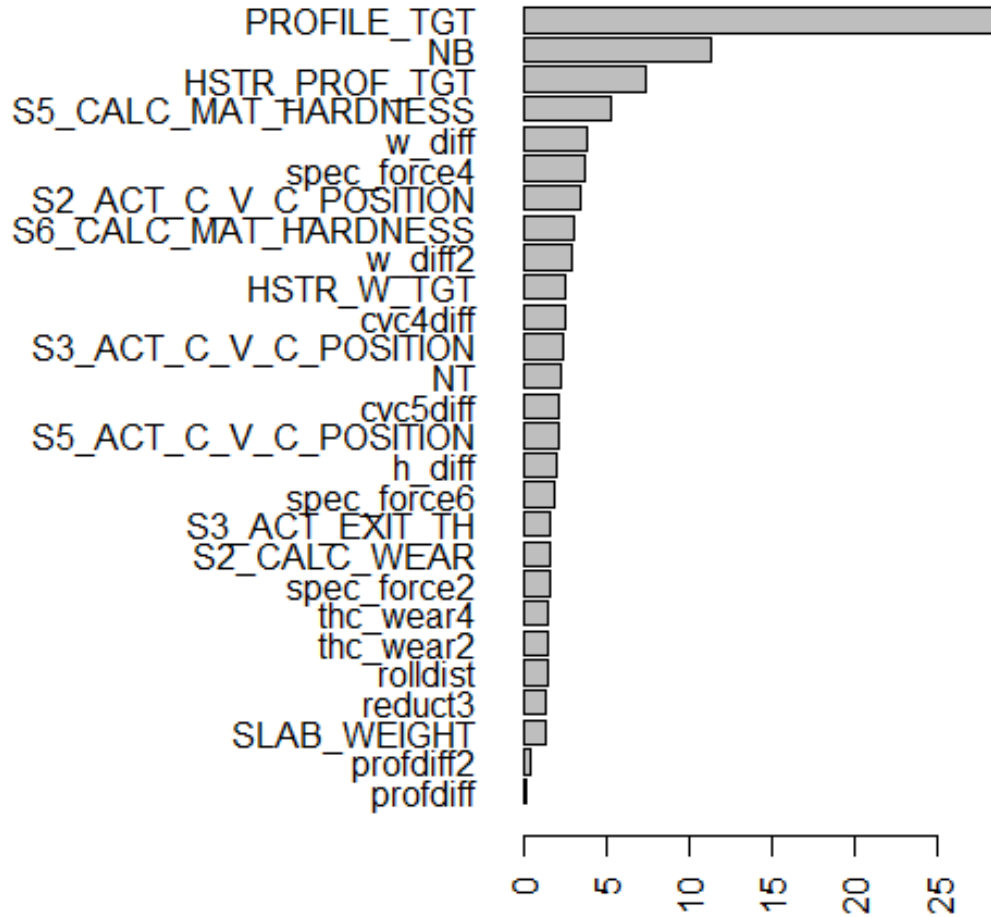
Opi enemmän ilmiöistä ennustemallien avulla



- Datapohjaisia malleja terästuotteen laadun ennustamiseen (SSAB, Outokumpu ja Ovako)
 - Erilaisia mekaanisia ominaisuuksia
 - Profiili
 - Tasomaisuus
 - Karheus
 - Keskilinjan paikka
 - Lämpötila
- Malleista voidaan johtaa myös riskitodennäköisyyden ennusteita
- Monimutkaisten teollisuusprosessien mallintaminen vaatii tehokkaita koneoppimismenetelmiä

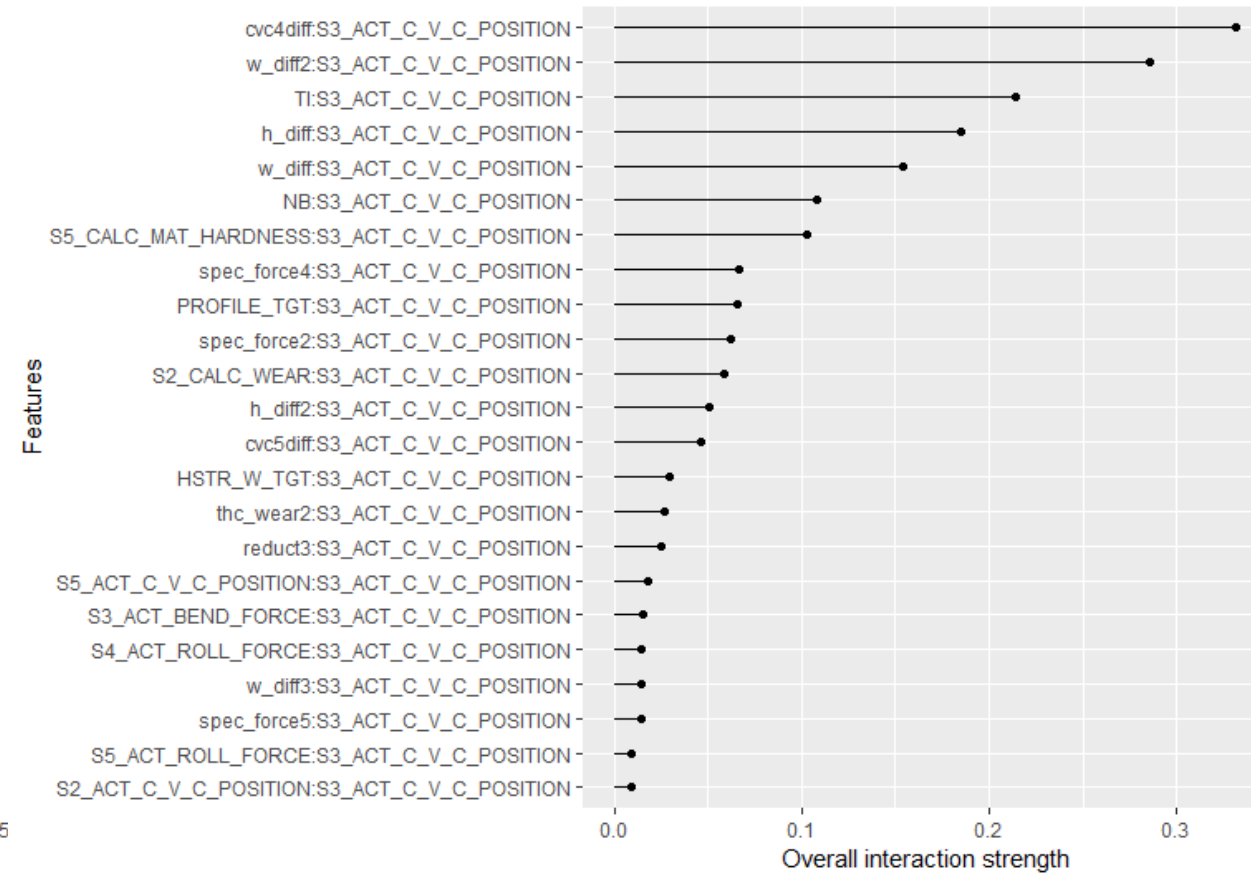
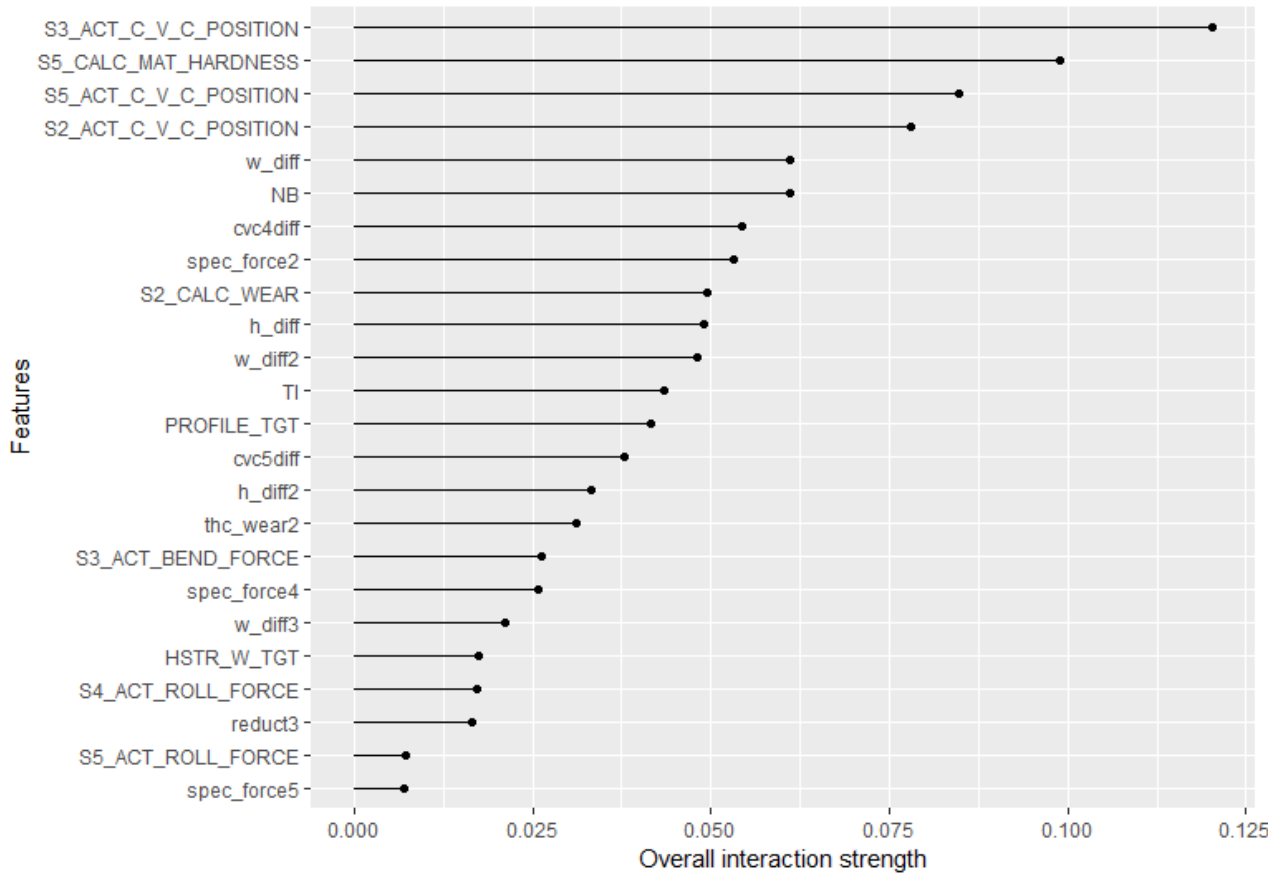


Valoa mustaan laatikkoon





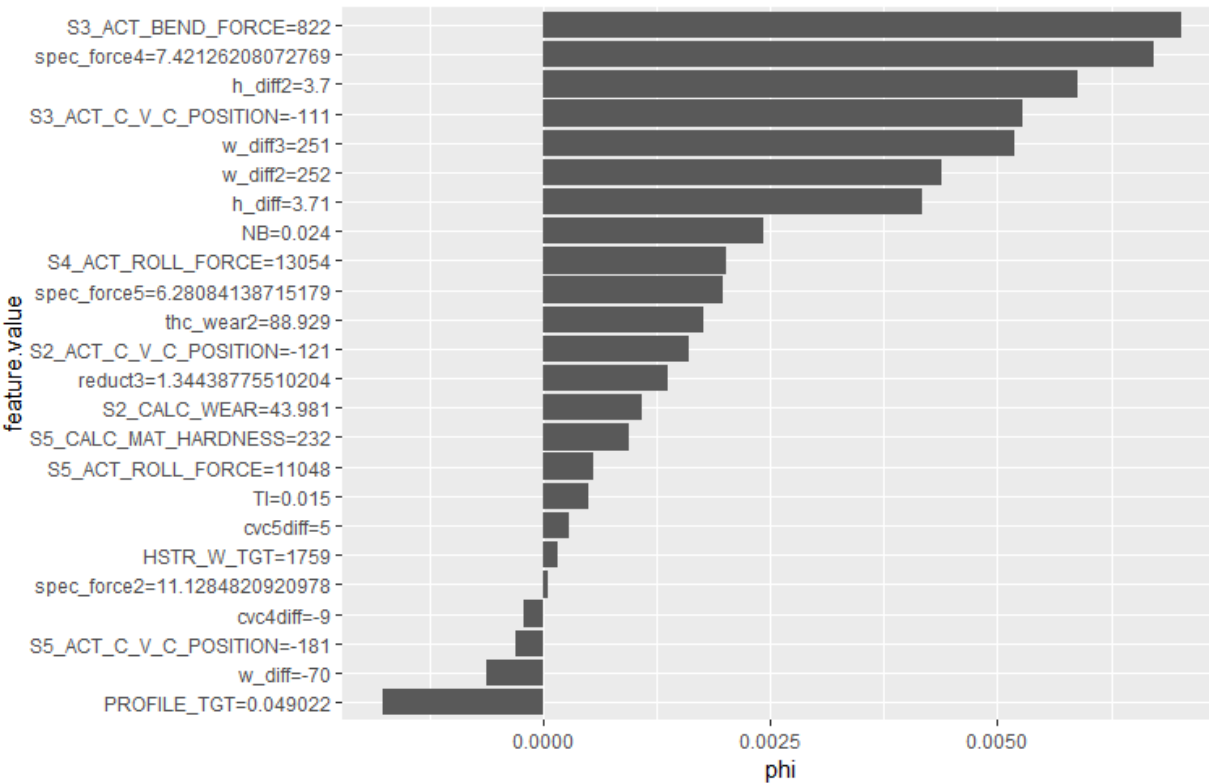
Interaktiot



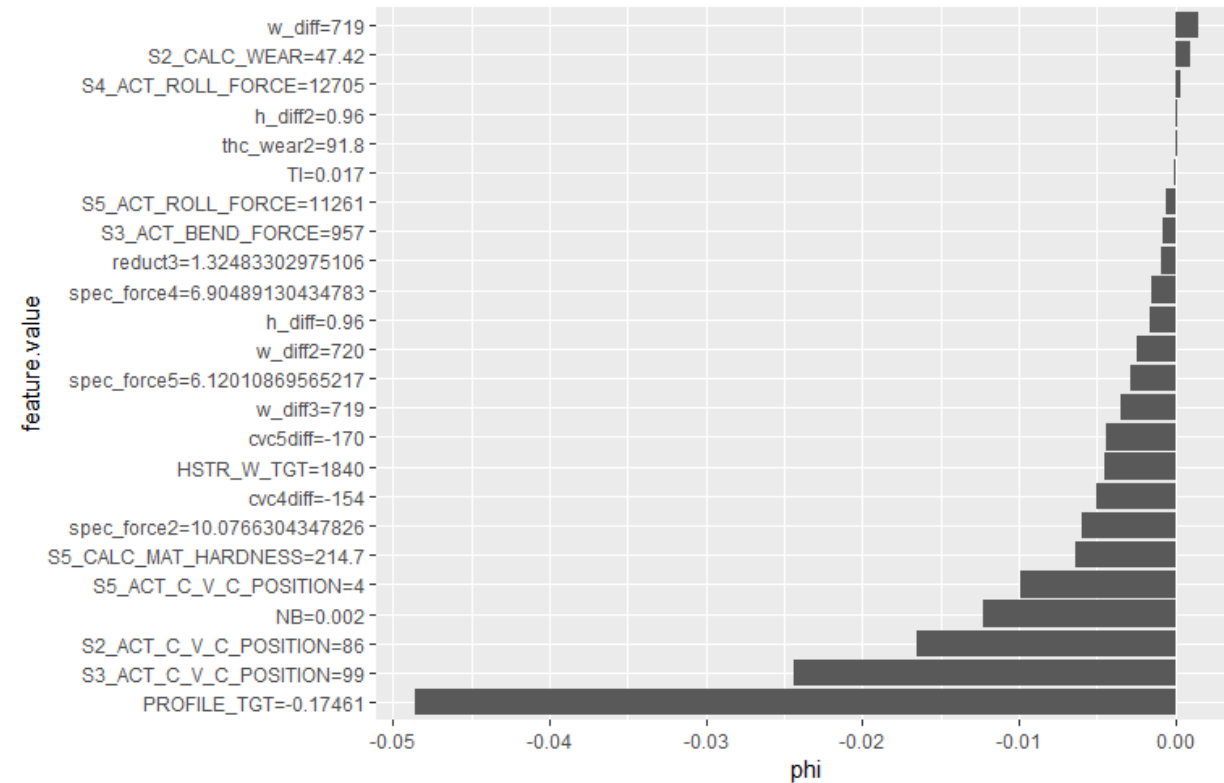


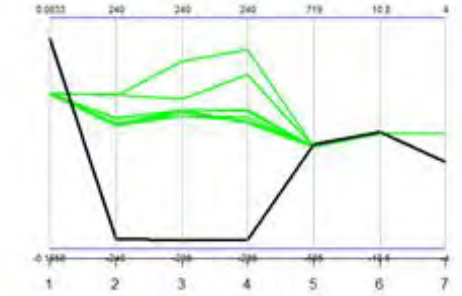
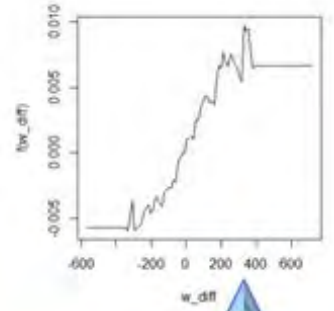
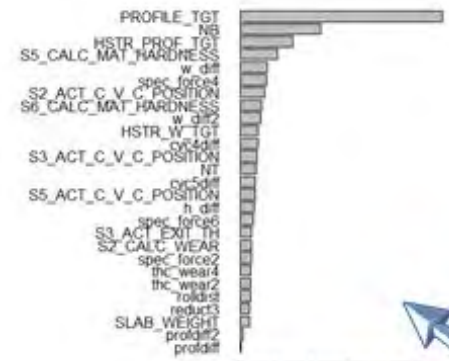
Mihin ennuste perustuu?

Actual prediction: 0.10
Average prediction: 0.05



Actual prediction: -0.10
Average prediction: 0.05

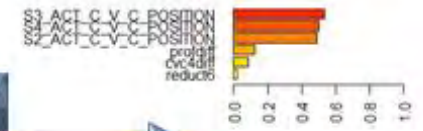




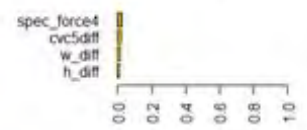
Quality during 16.1.2017 10:30 - 14:30. Number of strips: 48

Row	Color 1	Color 2	Color 3	Color 4	Dropdown
1	Red	Orange	Yellow	Green	cstkg
2	Red	Orange	Yellow	Green	sso prof
3	Red	Orange	Yellow	Green	prof ac

INCREASE VALUE:



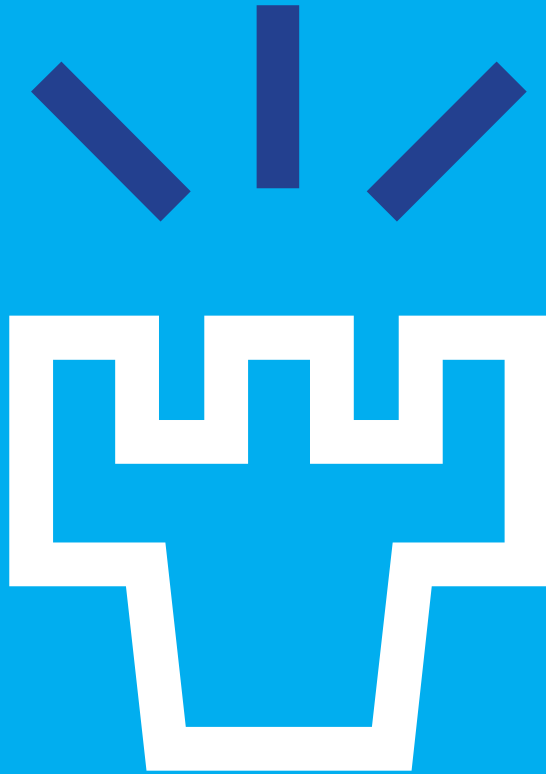
DECREASE VALUE:





Kokemuksiamme

- Ennustemallit voivat muodostaa herätteen tekoälylle, kun halutaan ennakoida
- Ennustemallit tuottavat tietoa myös mallinnettavasta prosessista
- Tarvitaan historiadataa, mielellään kaikista mahdollisista toimintatilanteista
- Dynaamiset ja oppivat mallit voivat mukautua muuttuviin tilanteisiin tai ne voidaan personoida yksilöllisiksi
- Mallille voidaan antaa palautetta, jolla opetusta tehostetaan
- Mallit voivat olla digitaalisen kaksosen pohjana



**UNIVERSITY
OF OULU**