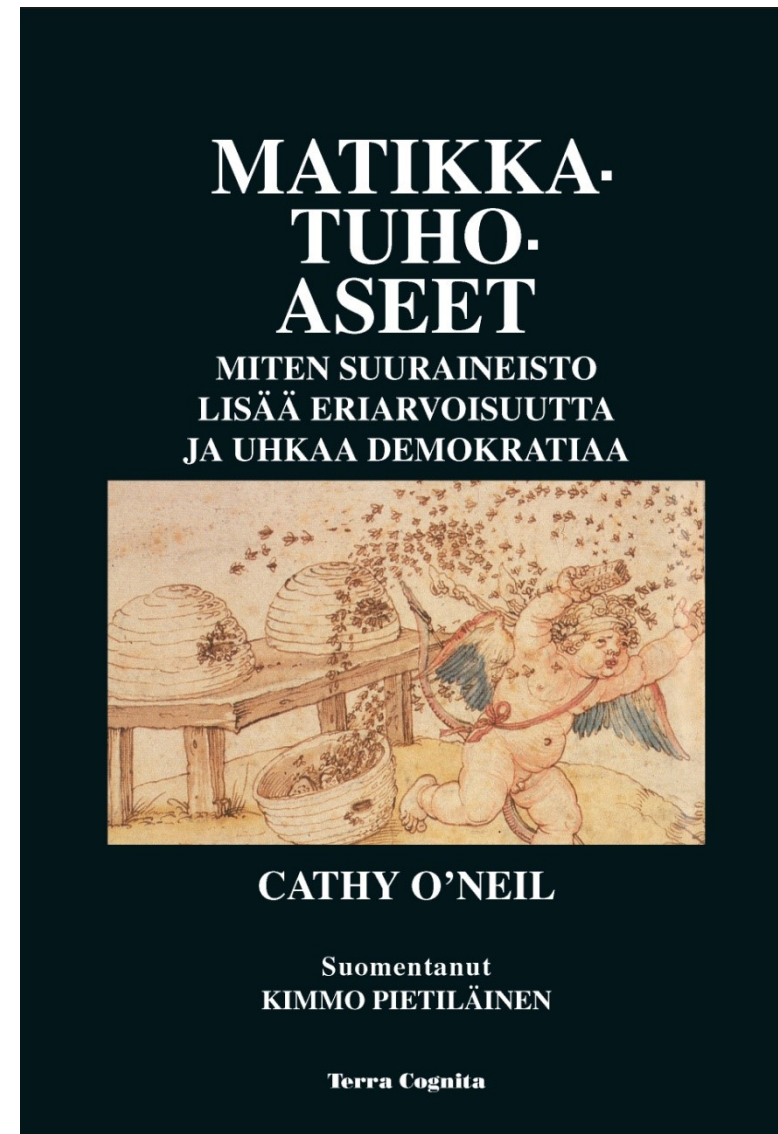


# Arviointialgoritmien vaaroja

Ilkka Norros

TEPin syysseminaari 23.11.2019

Cathy O'Neil: *Weapons of Math Destruction*.  
Crown, 2016.



# Automatisoitua arviointia työelämässä

- Rekrytointiin liittyvät seulonnat
  - Periaate: ”Tarkoitus ei ole löytää parasta, vaan karsia mahdollisimman monta mahdollisimman halvalla”
  - Käytännössä hakijan on laadittava ansioluettelonsa koneelle
  - Esimerkkitapaus: oppivalle ohjelmistolle syötettiin ihmisten päätöksiä
    - oppi harjoittamaan erinomaisen tehokasta syrjintää
  - Ihmisen oveluus ylittää yhä koneen
    - nokkelimmat keksivät usein, miten vähemmänkin pätevä pääsee pujahtamaan koneen suorittaman karsinnan läpi jatkoon
- Työn aikataulutus
  - Perustuu perinteikkääseen operaatioanalyysiin
  - Yhdistämällä optimointi sensoreilla tapahtuvaan tosiaikaiseen suoritusten mittaukseen voidaan työntekijöiden ajan ”tuhlausta” valvoa ja sanktioida äärimmäisen tarkasti

# Automatisoitua arviointia opetuksessa

Kun suuraineistoa tulee saataville, algoritmeilla tehdyt arvioinnit ja luokitukset yleistyvät

- Esimerkkitapaus:
  - kouluopettaja erotettiin oppilaiden huonojen arvosanojen vuoksi ottamatta huomioon oppilaiden vaihtumista
- Collegejen ja yliopistojen pisteyttäminen
  - US News & World Report –lehti keksi vuonna 1983
  - tutk.prof. Niilo Kauppi: Kiina globalisoi tämän ad hoc järjestelmän koptoimalla sen säännöt ns. Shanghain rankingiin
  - otetaan Suomenkin yliopistomaailmassa yhä enemmän tosissaan
    - absurdi arviointi muuttuukin toimintaa suuntaavaksi tekijäksi!

# Mainontaa heikkona hetkenä

Mainonnan kohdentaminen on aineistoanalyysin keskeisiä sovelluksia

- Tilanne ei ole mustavalkoinen:
  - moni haluaa saada mainoksia, jotka häntä kiinnostavat
  - moni toivoo, ettei häntä häirittäisi epäkiinnostavilla asioilla
- Eettisesti ongelmallista on ihmisten kokonaisvaltainen profilointi verkkokäyttäjien perusteella
  - koneoppivat järjestelmät alkavat ”tietää, mitä sanat merkitsevät”
  - ne pystyvät ennustamaan ihmisen käyttäytymistä
    - sisällöltään yksilöllisesti kohdennettua mainontaa
    - ovelasti ajoitettua mainontaa
  - matikkatuhoaseen kategorioita: tämän soveltaminen pikaluottojen tarjoamiseen vähävaraisille

# Rikoksia ennustamalla rikoksia

Matikkatuhoaseita esiintyy myös poliisin ja oikeuslaitoksen piirissä

- Yhtenä esimerkkinä PredPol-yhtiön järjestelmä
  - ennustaa rikosten tapahtumapaikkoja aineiston perusteella
  - käytetty poliisipartioinnin suunnitteluun joissakin USA:n kaupungeissa
  - laadittu muodollisesti neutraaliksi ja ennakkoluulottomaksi
  - ennustaa parhaiten vähäpätöisiä ”riesarikoksia”
  - piirre vahvistuu, kun partioinnin tuloksia syötetään oppivaan järjestelmään uutena aineistona
- O’Neil: riesarikollisuuden kartta kuvaa itse asiassa köyhyyttä
  - PredPolin tyyppiset järjestelmät ovat tehokkaampia köyhyyden kriminalisoimisessa kuin vakavien rikosten ehkäisemisessä

# Tahattomien vääristymien syitä

- Korvikemuuttujien käyttö
  - usein tavoiteltua ominaisuutta ei pystytä suoraan mittaamaan tai ennustamaan (esim. menestyminen työtehtävässä)
  - sen uskotaan kuitenkin ilmenevän joissakin muissa, mitattavissa suureissa (esim. persoonallisuustestin tulokset)
- Mallintajien valitsemat optimointitavoitteet
  - tehokkuus ja tuottavuus yleisiä tavoitteita
  - oikeudenmukaisuus ja reiluus voivat jäädä huomiotta
- Järjestelmät eivät useinkaan kerää palauteaineistoa
  - esim. työnhakijoita luokittelevalla järjestelmällä ei yleensä ole aavistustakaan siitä, kuinka monta erinomaista työntekijää se on hylännyt

# Ammattietiikka muotoutumassa?

- IEEE julkaisi tänä vuonna raportin [Ethically Aligned Design](#)
  - IEEE Standards Association'in tuotos
  - perusteellisesti valmisteltu, 294 s
- TEMin hanke Tekoälyaika (2017-19):
  - 70 yritystä ja laitosta sitoutui tekoälyn eettiseen hyödyntämiseen
- EU:n tietosuoja-asetus suojaa henkilötietoja, kieltää profilointia, vaatii läpinäkyvyyttä ym
- Mutta kulkeeko etiikka aina innovaatioiden jäljessä?