

Workshop
'Big Data': kansen en knelpunten
Festival van het Leren
16 september 2016

PRANO GROUP



Prano Group
kennis in organisatie

© Thieu Kuys, 2016

Workshop Big Data: festival van het leren, 16 september 2016

De vragen in deze workshop

Wat is 'Big Data' voor iets?

Welke kansen en mogelijkheden biedt 'Big Data'?

Wat zijn de risico's met 'Big Data'?

Welke gevolgen heeft 'Big Data' voor de (volwassenen)educatie?

Big Data: Enkele voorbeelden

- Human Genome Project**: erfelijke ziektes opsporen met DNA-kaart
- Google Flu Trends**: zoekmachine gebruiken om griepverspreiding in kaart te brengen
- WalMart** (Netflix, bol.com enz.): consumentenkeuzes volgen om consumentenvoorkeuren in kaart te brengen
- CBS**: inflatiecijfer a.h.v. kasregistratiesystemen
- CBS**: sociale media gebruiken om consumentenvertrouwen te peilen
- Google** (Tesla enz): zelfrijdende auto's (met zelfherstellende verf)
- Iedereen (en alles)**: 'Internet Of Things' / 'Internet of Everything'

Big Data: Enkele voorbeelden in het onderwijs

- Plagiaatherkenning** (Turnitin e.a.): plagiaat opsporen in scripties
- Learning Analytics** (div.): zoveel mogelijk gegevens over studieproces verzamelen om factoren van studiesucces in kaart brengen
- Educatieve software** (Oefenweb, Snappet): leerlingdifferentiatie zonder tussenkomst van docenten
- OCW (en alle scholen)**: onderwijs verbeteren door leerlingen te volgen ('BRON', leerlingvolgsystemen, leeropbrengst-metingen enz.)

Big Data = Gegevensverwerking

VERZAMELEN

ANALYSEREN

RESULTAAT

‘Big Data’ staat voor een bepaalde manier van gegevensverwerking of een bepaalde verwachting daarbij

Gegevensverwerking: verzamelen

VERZAMELEN

‘Klassiek’	‘Big Data’
Bepaalde soort gegevens	Verschillende soorten gegevens
Beperkte steekproef	Grote hoeveelheid, soms alle data
Voor een bepaald doel	Ongericht, soms bij toeval (bijvangst, secundaire data, meta-data): hergebruik van gegevens die voor een ander doel zijn verzameld

Gegevensverwerking: analyseren

ANALYSEREN

‘Klassiek’	‘Big Data’
Veel aandacht voor data-validatie, uitfilteren van ‘uitschieters’	Toestaan van ‘onjuiste’ data en ‘uitschieters’
Trage, ‘off-line’ analyse	Snelle, ‘on-line’ analyse
Geprepareerde algoritmes	‘Machine learning’: ‘zelflerende’ algoritmes

Gegevensverwerking: resultaat

RESULTAAT

‘Klassiek’	‘Big Data’
Hypothese-toetsing	Hypothese-vorming
Theorie → Advies	Advies → ‘actionable knowledge’
vaak theoretisch doel (niet commercieel)	Vaak praktisch doel (commercieel)

Big Data = hernieuwd geloof in twee idealen

MEETBAARHEIDS-IDEAAL

```
graph TD; A[MEETBAARHEIDS-IDEAAL] --> B["We kunnen zowat alles kwantificeren, meten"]; C[MAAKBAARHEIDS-IDEAAL] --> D["We kunnen zowat alles controleren en aanpassen"];
```

“We kunnen zowat alles kwantificeren, meten”

MAAKBAARHEIDS-IDEAAL

“We kunnen zowat alles controleren en aanpassen”

Kansen en mogelijkheden

Kloof tussen theorie en praktijk verkleind

Big Data-oplossingen werken soms (acceptabele ‘trade-off’ tussen betrouwbaarheid en snelheid)

Stereotiepen en dogma’s komen onder druk: uitzonderingen worden zichtbaar

Machtsstructuren kunnen worden doorbroken

Maar, natuurlijk, *“elk voordeel heb z’n nadeel”*...

Risico's en gevaren

‘Privacy’-risico's: huidige wetgeving inadequaar
(en nieuwe EU-regels ook)

‘Profiling’: uitzonderingen worden zichtbaar(der) maar dat
kan stereotiepen en dogma's ook versterken

Naarmate de kloof tussen theorie en praktijk kleiner wordt,
wordt de positie van het bedrijfsleven, de ‘markt’ sterker

De ‘politiek’ zal niet achter willen blijven

➔ ‘machts-wedloop’ gedreven door ‘controle-zucht’

En dan nog iets...

In de ban van de illusie – deel 1



59 kB

Welke foto bevat de
meeste informatie?

557 kB

In de ban van de illusie – deel 2



Welke informatie
geeft deze foto?



En deze foto?



In de ban van de illusie – deel 3

De ‘profiling’-paradox

DS = beslissing (vul maar wat in)

T_{DS} = beslissing DS is terecht/juist/redelijk/billijk

E = bewijsmateriaal

R = Risico

$P(X|Y)$ = Waarschijnlijkheid van X gegeven Y

ND = No Data

SD = Small Data

BD = Big Data

$$P(T_{DS}|E_{ND}) < P(T_{DS}|E_{SD}) < P(T_{DS}|E_{BD})$$

maar toch:

$$R(DS|BD) > R(DS|ND)$$

Hoe zit dat???

In de ban van de illusie – deel 4

Waar ‘Big Data’ succesvol is, is dat meestal in een ‘klassieke’ setting.

Door (soms niet begrepen) successen zijn de verwachtingen rond ‘Big Data’ overspannen geraakt.

Ondertussen is het fenomeen ‘Big Data’ is verweven geraakt met andere thema’s (permanente educatie, ’21st skills)

Gevolgen voor (volwassenen)educatie

Stelling 1: Mensen die op latere leeftijd gaan studeren hebben **MEER** kans van slagen: er zijn veel middelen beschikbaar om ‘achterstanden’ in te lopen.

Stelling 2: ‘Profiling’, ‘controlezucht’ en de overspannen verwachtingen rond ICT en ‘Big Data’ hebben tot gevolg dat de kansen voor deze groepen juist **KLEINER** worden.

Colofon



Prano Group
kennis in organisatie

De PRANO GROUP organiseert lezingen, cursussen en bijeenkomsten de besturingsfilosofie van organisaties verdiepen. PRANO is onafhankelijk en ongebonden en kan ontwikkelingen in sectoren met gepaste reflexieve distantie bespreken. PRANO werkt samen met deskundige adviseurs die sectoren van binnen en buiten kennen.

www.prano.nl

Thieu Kuys (1961) is onafhankelijk onderzoeker en zelfstandig ondernemer. Hij studeerde Nederlands MO en Geschiedenis MO aan de Katholieke Leergangen in Tilburg en Wijsbegeerte aan de Katholieke Universiteit Brabant. In 1992 promoveerde hij op een onderzoek naar de mogelijkheid om filosofische problemen op te lossen met wiskundige en empirisch-wetenschappelijke methoden. In 1997 raakte hij werkzaam in de 'ICT'. In 2000 startte hij een eigen 'ICT'-onderneming, CSGC Advisering. Hij schreef artikelen en boeken over filosofie en ICT en is momenteel bezig met de voorbereiding van een boek over 'Big Data in het onderwijs'.



Workshop Big Data: festival van het leren, 16 september 2016