

Political Data Science

Lektion 13: Refleksioner om data science

Undervist af Jesper Svejgaard, foråret 2018
Institut for Statskundskab, Københavns Universitet
github.com/jespersvejgaard/PDS

I dag

1. Announcement
2. Opsamling fra sidst
3. Dagens pensum
 - Case: Maskinlæring i en administrativ praksis
 - Metodologiske, epistemiske og normative problemstillinger
4. Case-analyse
5. Eksamens-workshoppen i næste uge
6. Afrunding og sidste gang

Overblik

1. Intro til kurset og R
2. R Workshop I: Explore
3. R Workshop II: Import, tidy, transform
4. R Workshop III: Programmering & Git
5. Web scraping & API
6. Tekst som data
7. Visualisering
8. GIS & spatiale data
9. Estimation & prædiktion
10. Superviseret læring I
11. Superviseret læring II
12. Usuperviseret læring
13. Refleksioner om data science
14. Opsamling og eksamen

Announcement

Speciale-arrangement

Speciale-arrangement i Digital Political Science Group d. 23. maj kl. 15 - 16.

Afholdes i 2.0.30.

Mere info kan findes [her](#)

Opsamling fra sidst

Opsamling fra sidst

Undervisning og øvelser

Undervisning:

- PCA
 - Reduktion af dimensioner som præprocessering
 - Hvilke dimensioner i datasættet er mest interessant?
 - Hvordan hænger de uafhængige variable sammen?
- K-means clustering:
 - Gruppering af data, når vi ikke har et label på forhånd
- Hierarkisk clustering:
 - Gruppering af data - uden præspecificering af antal clusters
 - Dendrogrammer

Motivation

Motivation

Hvorfor refleksion om data science?

To helt essentielle problemstillinger:

1. Hvordan går vi fra en AUC til praktisk anvendelse af en prædiktionsmodel?
2. Hvad sker der, når vi implementerer en prædiktionsmodel i en praksis?

Case: Maskinlæring i en administrativ praksis

Case: Metropol

Casen kort

Koncept:

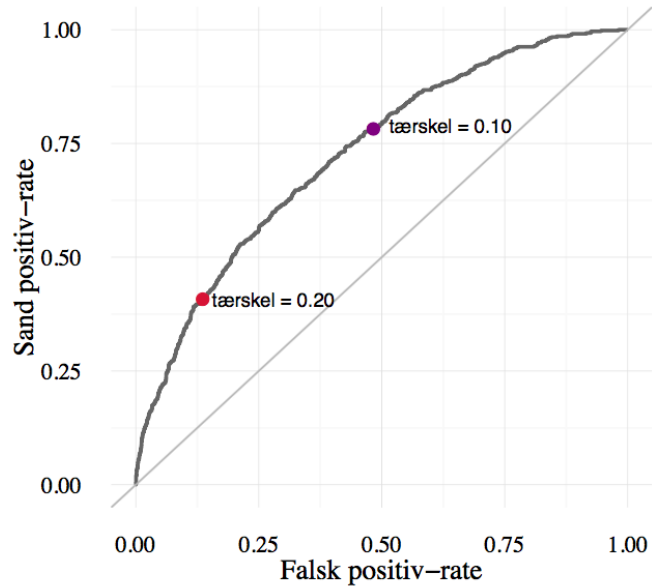
- Har ML potentiale til målretning af tiltag mod uddannelsesfrafald?
- Frafald inden for første studieår ved studiestart blandt tidligere studerende

Anvendelse af data science i praksis:

- Hvor præcist kan vi forudsige uddannelsesfrafald?
- Hvordan omsætter vi en prædiktionsmodel til handling?
- Hvad er konsekvenserne af målretning af tiltag pba. ML?

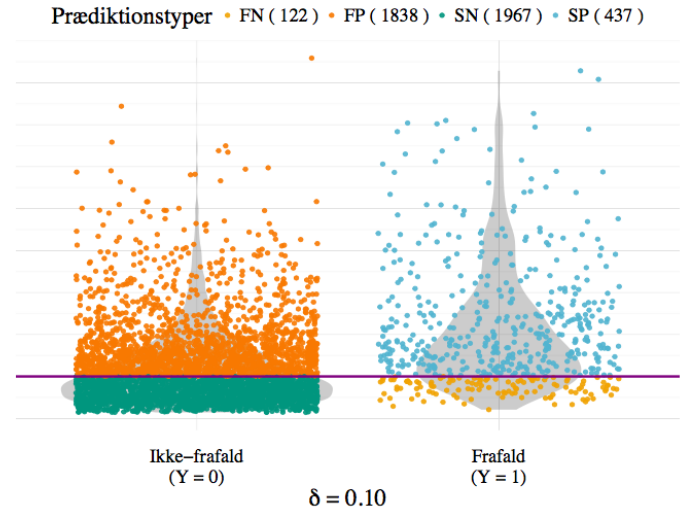
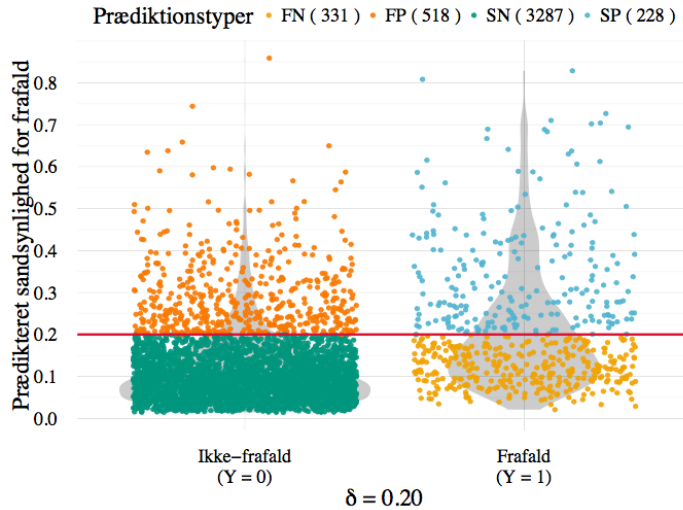
Case: Metropol

Måretning og tærskelværdi



Case: Metropol

Betydningen af tærskelværdi I



Case: Metropol

Betydningen af tærskelværdi II

| $\delta = 0.20$ | | | | $\delta = 0.10$ | | | |
|-----------------|-----------------|-----------------|-------------|-----------------|-----------------|-----------------|-------------|
| | $\hat{y}_i = 1$ | $\hat{y}_i = 0$ | | | $\hat{y}_i = 1$ | $\hat{y}_i = 0$ | |
| $y_i = 1$ | 228 | 331 | SPR = 0,408 | $y_i = 1$ | 437 | 122 | SPR = 0,782 |
| $y_i = 0$ | 518 | 3287 | FPR = 0,136 | $y_i = 0$ | 1838 | 1967 | FPR = 0,483 |

Case: Metropol

Tiltag

| | Mail | Visitationsmøde | Mentorforløb |
|---|----------------|--------------------|---------------------|
| Omkostning i kr. | 4200 kr. i alt | 420 kr. per tilbud | 4200 kr. per tilbud |
| Omkostning i timer | 10 timer i alt | 1 time per tilbud | 10 timer per tilbud |
| Antaget gennemsnitlig fastholdelseeffekt | 0,2 % | 2 % | 12,5 % |

Case: Metropol

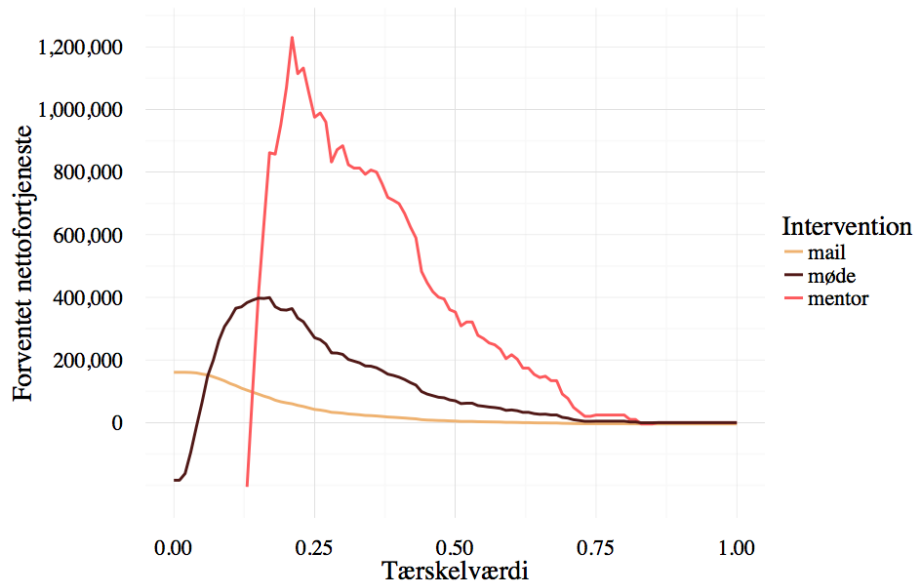
Cost-benefit: Framework til målretning

Forventet nettofortjeneste:

$$\begin{aligned} F - O &= SP \times \alpha \times \text{indtægtssats} - (SP + FP) \times \text{udgiftssats} \\ &= 228 \times 0,02 \times 147.486 - (228 + 518) \times 420 \\ &= 359.216,16 \end{aligned}$$

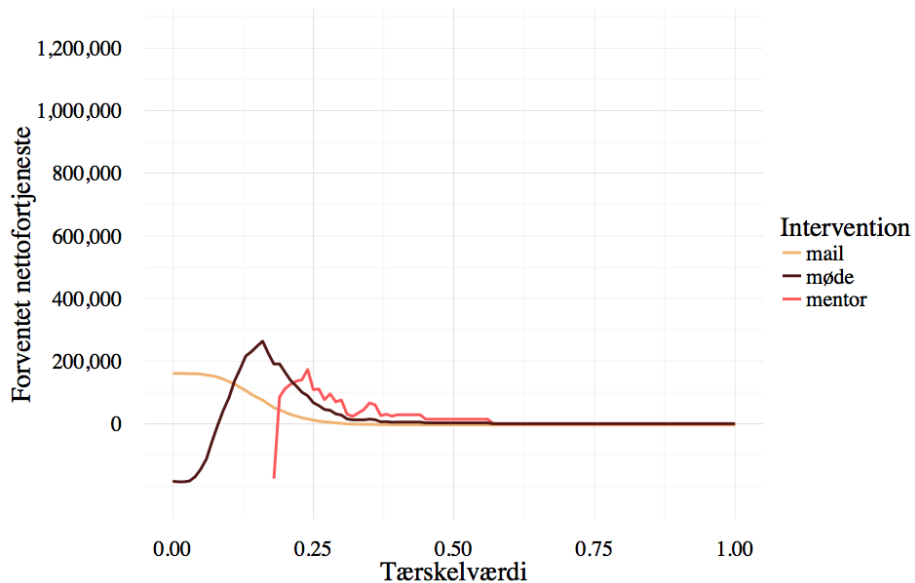
Case: Metropol

Målbetning med GBT



Case: Metropol

Målbretning med logit



Case: Metropol

Resultater

Indsigter:

- Vi kan nu fastsætte den optimale tærskelværdi empirisk
- Vi kan fastsætte tærskelværdien afhængigt af politisk målsætning
- Prædiktionsperformance har stor betydning i praksis

Metodologiske problemstillinger

Metodologiske problemstillinger

Overblik

Centrale methodologiske problemstillinger:

- Teoriens rolle
- Estimation vs. prædiktion (genbesøgt)
- Prædiktions over tid

Metodologiske problemstillinger

Teoriens rolle I

Anderson (2008):

- "Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With **enough data**, the numbers **speak for themselves**."
- "... faced with massive data, this approach to science - hypothesize, model, test - is becoming **obsolete**."
- "There's no reason to cling to our **old ways**. It's time to ask: What can science learn from Google?"

Metodologiske problemstillinger

Teoriens rolle II

Anderson (2008):

- Vi kan roligt bare lade være med at bekymre os om frafaldslitteraturen. Rart!

Mittelstadt et al. (2016):

- "... statistical methods can help **identify significant correlations**, but these are rarely considered to be sufficient to posit the existence of a causal connection and thus may be **insufficient to motivate action**."

Eksempel:

- Vi ved kun, hvem der dropper ud, ikke hvorfor de gør det - hvordan skruer vi så et godt tiltag sammen?

Metodologiske problemstillinger

Estimation vs. prædiktions genbesøg I

Anderson (2008):

- "Correlation is enough. We can stop looking for models."
- "Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all."

Metodologiske problemstillinger

Estimation vs. prædiktions genbesøg II

Anderson (2008)

- Bekymringer om spuriøsitet, kollinearitet og selektionsbias er forældede!

Lazer et al. (2014):

- "**Big data hubris** is the often implicit assumption that big data are a substitute for, rather than a supplement to, traditional data collection and analysis."
- "... quantity of data does not mean that one can ignore **fundamental issues** of measurement and construct validity and reliability and dependencies among data."

Hvilke fundamental issues har betydning for vores case om uddannelsesfrafald?

Metodologiske problemstillinger

Estimation vs. prædiktion genbesøgt III

Samii (2016)

- **Post treatment-bias:** Inklusion af variable der kan følge af treatment
- Eksempel: Tiltag mod frafald pba. fx lave karakterer

Athey (2017)

- **Heterogene kausale effekter:** Når effekten af tiltag varierer
- Eksempel: Reagerer de mest/mindst frafaldstruede ens på treatment?

Metodologiske problemstillinger

Estimation vs. prædiktions genbesøg IV

"Overall, for big data to achieve its full potential in business, science, and policy, **multidisciplinary approaches** are needed that build on new computational algorithms from the SML literature, but also that bring in the methods and practical **learning from decades** of multidisciplinary research using **empirical evidence** to inform policy." (Athey, 2017)

=> Hybrid-tilgang mellem præd. og est. - og måske nye problemstillinger?

Metodologiske problemstillinger

Prædiktion og målretning over tid

Stabilitet (Athey, 2017)

- En retvisende model kræver et stabilt miljø
- Minder om Lucas-kritikken: Statistiske modeller er trænet på historiske data
- Hvilken betydning får fremdriftsreformen eller ændringer i SU'en?

Algorithm dynamics (Lazer et al., 2014)

- Hvad sker der med en prædiktionsmodel over tid?
- Har tiltagene en effekt => ændrer på korrelationer ml. variable og outcome
- Minder om problemet med Google Flu Trends: Algoritmen føder ind i sig selv

Havde vi været bedre stillede, hvis vi havde grebet problemet an som et estimationsproblem?

Epistemiske problemstillinger

Epistemiske problemstillinger

Overblik

Mittelstadt et al. (2016):

- Inconclusive evidence
- Inscrutable evidence
- Misguided evidence

Epistemiske problemstillinger

Inconclusive evidence I

"When algorithms draw conclusions from the data they process using inferential statistics and/or machine learning techniques, they produce **probably** yet inevitably **uncertain knowledge**" (Mittelstadt et al., 2016)

Udfordring:

- "All models are wrong, but some are useful" - også i en adm. praksis?
- 70 % for frafald = 30 % for ikke-frafald.
- Hvis den studerende ikke frafalder, var modellen så forkert?
- Hvor sikre skal vi være på en prædiktion for at vi kan handle på den?

Epistemiske problemstillinger

Inconclusive evidence II

Udfordringer (Hofman et al., 2017):

- Ingen standarder
 - researcher degrees of freedom => falske positive
 - genbrug af out-of-sample data => "human-in-the-loop overfitting"
- Grænser for prædiktions-performance
 - Hvor forudsigelig er den menneskelige natur?
 - Reducerbare vs. ikke-reducerbare fejl
- Prædiktionsperformance vs. forståelighed
 - Afvejning ml. modelkompleksitet og forståelighed

Epistemiske problemstillinger

Inconclusive evidence III

Løsninger (Hofman et al., 2017):

- Ingen standarder
 - Kig til AI-litteraturen => standardiserede benchmarks
 - Skeln ml. forskningstyper, fx eksplorativ og hypotesetestende
- Grænser for prædiktions-performance
 - Forhold dig til **mulig** baseline performance
 - Outcome givet af random variable? (teoriens rolle)
- Prædiktionsperformance vs. forståelighed
 - Simple modeller ikke nødvendigvis mere generaliserbare (researcher DF)
 - Komplekse modeller performer ikke nødvendigvis bedre
 - Forstå kan både være "sensemaking" og "forklaring"

=> hybrid-tilgang: fastlæg design på forhånd, åbenhed, klarhed omkring mål

Epistemiske problemstillinger

Inscrutable evidence

Mittelstadt et al. (2016):

- "When data are used as evidence for a conclusion, it is reasonable to expect that **the connection between the data and the conclusion** should be **accessible**."
- "Opacity in machine learning algorithms is a product of the high dimensionality of data, complex code and changeable decision making logic."
- "The rationale of an algorithm can in contrast be **incomprehensible to humans**, rendering the **legitimacy** of decisions difficult to challenge."

Konsekvenser:

- Tab af tillid til beslutninger truffet i black boxes
- Opprioritering af simple algoritmer?

Epistemiske problemstillinger

Misguided evidence I

Mittelstadt et al. (2016):

- "Algorithms process data and are therefore subject to a limitation shared by all types of data-processing, namely that **the output can never exceed the input.**"
- "Algorithms inevitably make **biased decisions.**"

Epistemiske problemstillinger

Misguided evidence II

Kilder til bias:

1. Præ-eksisterende værdier

- Data genereres ikke af sig selv - data findes ikke bare.
- Hvad måler vi? Hvordan måler vi det? Systematiske fejl i målingen?
- Fx typisk nemmere at måle antal end kvalitet

2. Tekniske begrænsninger

- Forskellige algoritmer giver forskellige prædiktioner
- Hvordan sætter vi de frie parametre for fx validering?
- Hvor langt tilbage bruger vi data fra? Alle variable? Imputation?

3. Emergerende aspekter

- Hvilke problemer og løsninger indgår i modellen?
- Hvilke treatments målretter vi?
- Hvilket outcome målrette vi imod?

Epistemiske problemstillinger

Misguided evidence III

Ville vi ikke være bedre stillet med en menneskelig sagsbehandler?

Normative problemstillinger

Normative problemstillinger

Overblik

Ordnet efter Mittelstadt et al. (2016):

- Traceability / accountability
- Unfair outcomes og diskrimination
- Transformative effekter

Normative problemstillinger

Accountability

Accountability:

- "This implies that **harm caused** by algorithmic activity is hard to debug ... [and] it is rarely straightforward to identify **who should be held responsible** for the harm caused." (Mittelstadt et al., 2016)

Hvem skal holdes ansvarlig?

- Programmøren? Dataindsamleren? Beslutningstageren?
- Hvad hvis algoritmen er selvlærende?
- Kan man give algoritmen et ansvar?
- Er det ønskværdigt?

Tænk på casen Metropol. Hvem bør holdes ansvarlig?

Normative problemstillinger

Unfair outcomes og diskrimination I

Mittelstadt et al. (2016):

- "... algorithms identify correlations and make **predictions about behaviour at a group-level** ... the individual is comprehended based on connections with others identified by the algorithm, rather than **actual behaviour**."
- "Even if strong correlations or causal knowledge are found this knowledge may only **concern populations** while **actions are directed towards individuals**"

Forskelsbehandling:

- Forskelsbehandling = en præmis for målretning af tiltag
- Forskelsbehandling behøver ikke være kontroversielt: Billetter, screening ...
- Hvor går grænsen? ... Forsikringspræmier? Prisen på en klipning?

Normative problemstillinger

Unfair outcomes og diskrimination II

Statistisk diskrimination:

- Er det rimeligt at behandle individer pba. gruppetilhørsforhold?
- Debatten vedrører som regel særlige kategorier som køn, etnicitet, religion
- Desuden skelnes ofte mellem positiv og negativ forskelsbehandling

Vil det være rimeligt at målrette tiltag mod studerende af anden etnisk herkomst end dansk, givet at de statistisk set har stor tendens til at frafalde?

Normative problemstillinger

Unfair outcomes og diskrimination III

Løsning 1:

- Vi fjerner sensitive variable såsom køn og etnicitet fra prædiktionsmodellen.

Løsning 2:

- Vi prøver at mætte modellen med alt tilgængeligt data - også køn, etnicitet mv.

Zoom ud:

- Hvad vil det egentlig sige at blive behandlet som individ frem for gruppe?
- Hvilken viden har vi, der ikke er baseret på tidligere erfaringer?

Hvad stiller os bedst - den menneskelige sagsbehandler eller algoritmen?

Normative problemstillinger

Transformative effekter I

Mittelstadt et al. (2016):

- "This is because algorithms can affect how we **conceptualise the world**, and modify its social and political organisation. Algorithmic activities, like profiling, **reontologise the world** by understanding and conceptualising it in new, unexpected ways, and triggering and motivating actions based on the insights it generates."

Paradoks:

- På den ene side kan individualiserede tiltag støtte den enkelte
- På den anden side begrænser målretningen af tiltag individets mulighedsrum

Normative problemstillinger

Transformative effekter II

Selvopfyldende profetier:

- Labels kan være selvopfyldende profetier.
- Fx: Hvilken betydning har det at få et label på sig som frafaldstruet?

Handle-muligheder:

- Individualiserede prædiktioner \neq individualiserede tiltag
- Fx: De uafhængige variable kollapses til to grupper

Identifikation af årsager:

- Vi kan kun se de kategorier, vi har konstrueret.
- Fx: Vi kan godt måle holdstørrelse som årsag til frafald - men hvad med dybereliggende strukturelle rødder?

Ved at ordne verden i variable og kategorier former vi, hvordan verden forstås og hvordan vi kan handle.

Opsamling

Opsamling

Framework til methodologiske, epistemiske og normative problemstillinger

Methodologiske problemstillinger:

- **Teoriens rolle:** Hvor vigtigt er teori?
- **Estimation vs. prædiktion:** Korrelationer og actionable insights
- **Prædiktationer over tid:** Stabilitet og algorithm dynamics

Epistemiske problemstillinger:

- **Inconclusive evidence:** Fejlbarlighed og åbenhed herom
- **Inscrutable evidence:** Transparens og forståelighed af algoritmer
- **Misguided evidence:** Præeksisterende værdier, teknisk bias, emergerende aspekter (garbage in/out)

Normative problemstillinger:

- **Traceability / accountability:** Hvem står til ansvar?
- **Unfair outcomes og diskrimination:** Behandling af gruppetilhørsforhold
- **Transformative effekter:** Reontologisering, selvopfyldende profetier, identifikation af årsager

Case-analyse

Case

Københavns Kommune overvejer at få nogle snarrådige scient.pol'ere til at udvikle en prædiktionsmodel, som skal implementeres i kommunens jobcentre. Modellen har til formål at forudsige hvilke ledige, der bliver langtidsledige, så jobcentrene kan målrette en præventiv indsats. Samtidig forventes modellen at have et effektiviserings-potentiale ved at kunne erstatte de sagsbehandlere, som ellers skulle være kontaktfladen mellem jobcentre og borgere uden arbejde.

Øvelse:

- Hvilke fordele og ulemper ser I ved casen? Brug evt. frameworket fra tidligere.
- Hvad vil I rådgive borgmesteren til - er ovenstående en god eller dårlig idé?
- **Diskuter** i mindre grupper med jeres sidemakkere.

Afrunding

Sidste lektion

Oplæg og opponering

Upload op til 1 sides oplæg om jeres seminaropgave til Canvas/Absalon under **Discussions**. Hver person skal uploade et oplæg. Oplægget kan indeholde jeres problemstilling, foreløbige tanker, jeres hovedudfordringer, idé til opgavens struktur - eller hvad I finder mest relevant at få feedback på. Oplæggene er udgangspunktet for oplæg/opponering i sidste lektion.

Proces:

1. Upload one-pager om jeres seminaropg. før **søn. d. 6. maj kl. 23.59**.
2. Husk at anføre **navn** og **hold** øverst i jeres oplæg!
3. **Mandag morgen** sender jeg en mail ud med fordeling i **grupper**.
4. **Læs oplæggene** i jeres gruppe inden undervisningen.

Sidste gang

- Indhold:
 - Eksamen og afrunding
- Forberedelse:
 - Upload oplæg senest **søndag d. 6. maj**
 - Læs gruppens oplæg før næste undervisning
- Pensum:
 - :)
- DataCamp:
 - :)

Tak for i dag!