

Datalogiens Videnskabsteori 2024

Ugeseddel 7

Indhold

Fælles info uge 7	1
Alle materialer til uge 7	1
Centrale begreber	2
Flipped Forelæsning 7 mandag den 3. juni	2
Klassisk forelæsning, mundtlig evaluering og spørgetime fredag den 7. juni	2
Øvelser uge 7	2
Emne 1: ML og bias: etik	3
Emne 2: Etik og professionelt ansvar	4

Fælles info uge 7

Der er ingen forelæsningvideoer eller nye grundbogskapitler til denne sidste uge i kurset. Der er stadig en Flipped forelæsning mandag den 3. juni, øvelser mandag/fredag eller tirsdag og en sidste klassisk forelæsning fredag den 7. juni.

Alle materialer til uge 7

Forelæsningsvideoer

n/a

Tilhørende kapitler fra grundbogen

n/a

Grundbogskapitel til klassisk forelæsning

n/a

Tekster til øvelserne

- Angwin et al. (2016)
- Friedler et al. (2021)
- Duhigg (2012)
- Gigerenzer (2017)
- (Sørensen and Johansen, 2021, kap. 10)
- ACM Code of Ethics (2018) (ikke pensum)
- The Tech Pledge (2019) (ikke pensum)

Centrale begreber

Der er ingen nye centrale begreber, siden der ikke er nogen videoer eller nye grundbogs-kapitler i denne uge.

Flipped Forelæsning 7 mandag den 3. juni

Til Flipped forelæsningen vil vi samle op på quiz-svarene fra uge 6 og lave aktiviteter om de centrale begreber indenfor etik og redelighed (gennemgået til de klassiske forelæsninger i uge 5 og 6).

Klassisk forelæsning, mundtlig evaluering og spørgetime fredag den 7. juni

Den klassiske forelæsning vil i den første time handle om studerendes akademiske redelighed og tvivlen om hvornår man må hvad ifølge reglerne; hvornår må man tale med sin medstuderende, hvornår må man bruge LLMs, etc. Mikkel har inden forelæsningen sendt et anonymt og frivilligt survey ud, som der bliver taget udgangspunkt i til forelæsningen.

I den resterende tid vil vi lave en kort mundtlig evaluering af kurset. Jeres input er meget brugbart for os og vi bruger aktivt jeres feedback til at forbedre kurset.

Vi vil også bruge noget tid iden sidste time som spørgetime, hvor I kan stille spørgsmål af formel eller indholdsmæssig karakter til kurset og til eksamen. Der vil ikke være andre spørgetimer inden eksamen. Hvis I stiller spørgsmål inden forelæsningen i diskussionstråden på Absalon, kan vi forberede os på det, I er mest interesserede i.

Øvelser uge 7

MANDAG 3. JUNI / FREDAG 7. JUNI ELLER TIRSDAG 4. JUNI

Emne 1: ML og bias: etik

Læringsmål

Efter dette emne skal du være i stand til at forklare begreberne *group fairness* og *individual fairness* samt forklare hvorfor de er gensidigt udelukkende. Du skal kende COMPAS-casen og kunne bruge den i en diskussion om udvikleres moralske og professionelle ansvar. Du skal derudover kende og kunne kritisk reflektere over Rybergs undskyldninger.

Litteratur (tilgængelig på Absalon)

- Angwin et al. (2016)
- Friedler et al. (2021)
- (Sørensen and Johansen, 2021, kap. 10)

Fremlæggelse gruppe 8: ML og bias - COMPAS-casen

1. Redegør for hovedtrækkene i sagen om COMPAS-algoritmen som den beskrives i Angwin et al. (2016).
2. COMPAS-algoritmen inddrager ikke direkte etnicitet, men den inddrager faktorer, der kan være korreleret med etnicitet (Angwin et al., 2016, 3–4). Har man som modelkonstruktør et ansvar for at undlade at inddrage faktorer, der kan korreleres med etnicitet eller andre følsomme forhold? Kan man det?

I artiklen Friedler m.fl. (2021) opbygges et meget detaljeret begrebsapparat, der kan bruges til at beskrive og forstå retfærdighed i maskinbeslutninger. I skal i fremlæggelsen bare tage udgangspunkt i de to centrale begreber.

3. Forklar begreberne *group fairness* og *individual fairness* fra Friedler et al. (2021) og forklar hvorfor en algoritme (normalt) ikke kan udvise gruppe- og individuel fairness samtidig.
4. Analyser hvordan disse fairness-begreber er på spil i COMPAS-casen.

Gruppearbejde emne 1

Diskuter i grupper følgende spørgsmål og hver klar til at præsentere jeres overvejelser i plenum:

1. Vurdér om følgende argumenter fritager udviklerne af COMPAS for et etisk ansvar:
 - (a) Vi har ikke gjort noget ulovligt.
 - (b) Vi udviklede systemet, men det var ikke os, der brugte det.

- (c) Vi gjorde bare, hvad der blev sagt.
 - (d) Hvis vi ikke havde udviklet systemet, havde nogle andre gjort det i stedet.
2. Er der et andet argument, der fritager udviklerne af COMPAS for et etisk ansvar, som I ville finde overbevisende?
 3. Diskuter hvorvidt udviklere bør forholde sig til hvorvidt de systemer, de udvikler, er etisk forsvarlige. Hvordan skal man fx. som udvikler beslutte hvilken form for fairness der indlejres i ens model?

Emne 2: Etik og professionelt ansvar

Læringsmål

Efter dette emne skal du være i stand til at forklare og bruge Thomas Nagels etiske ramme til diskutere de etiske komponenter af en case. Du skal også kunne kritisk reflektere over dit eget professionelle ansvar som medlem af faggruppen af tech-professionelle.

Litteratur

- Duhigg (2012)
- Gigerenzer (2017)
- (Sørensen and Johansen, 2021, kap. 10)

Fremlæggelse gruppe 9: Professionelt ansvar - Target-casen

1. Redegør i korte træk for Targets udvikling og brug af algoritmer til prædiktion af kundeadfærd, som den beskrives i Duhigg (2012).
2. Diskutér om Andrew Pole handlede etisk korrekt. Analysér de etiske komponenter i situationen ved at inddrage Thomas Nagels' skel mellem forskellige typer af etiske værdier (som det beskrives i Sørensen and Johansen, 2021, kap. 10).
3. Hvad kunne Andrew Pole have gjort for i højere grad at opføre sig etisk korrekt?

En del negative historier om bag data-analyse tager udgangspunkt i cases, hvor et firma benytter oplysninger om kunder eller brugere til at fremme deres egne kommercielle interesser. Man kan imidlertid også forestille sig, at de oplysninger, som firmaer som Google, Microsoft og Facebook sidder inden med, kan bruges til at udlede oplysninger, der vil være i brugernes interesse at kende til.

4. Diskutér på baggrund af Gigerenzer (2017) om Microsoft og Google har en etisk pligt til at udvikle algoritmer, der kan prædiktere væsentlige helbredsoplysninger om de mennesker, der benytter deres søgemaskiner, og at stille disse oplysninger til rådighed for, de brugere, der måtte ønske det.

Gruppearbejde emne 2

En måde at håndtere professionelt ansvar er strukturelt gennem opsætning af regler i fagfællesskaber. For eksempel har den internationale forening ACM (Association for Computing Machinery) har et *Code of Ethics* for sine medlemmer.

En anden version er det såkaldte TechPledge (The Tech Pledge, 2019), som blev udviklet på en Techfestival i København i 2019 af medlemmer af tænketanken "The 150". Ideen var at udvikle et generelt moralsk kodeks (en analogi til lægeløftet) for alle i tech-industrien.

Spørgsmål til diskussion

Overvej i grupper om I kan stå inde for de forskellige ansvarspunkter, som fremhæves i The Tech Pledge (listet nedenfor). Hvordan skal de revideres, så I ville skrive under på løftet? Er der noget, I synes der mangler?

The Tech Pledge siger følgende (The Tech Pledge, 2019):

I commit:

1. to take responsibility for what I create.
2. to only help create things I would want my loved ones to use.
3. to pause to consider all consequences of my work, intended as well as unintended.
4. to invite and act on criticism, even when it is painful.
5. to ask for help when I am uncertain if my work serves my community.
6. to always put humans before business, and to stand up against pressure to do otherwise, even at my own risk.
7. to never tolerate design for addiction, deception or control.
8. to help others understand and discuss the power and challenges of technology.
9. to participate in the democratic proces of regulating technology, even though it is difficult.
10. to fight for democracy and human rights, and to improve the institutions that protect them.
11. to work towards a more equal, inclusive and sustainable future for us all, following the United Nations global goals.
12. to always take care of what I start, and to fix what I break.

I take this pledge as an individual working to make tech a force for good in society, and I will challenge everyone I work with to do the same.

Litteratur

- ACM Code of Ethics (2018). ACM code of ethics. <https://ethics.acm.org/code-of-ethics/>. Association for Computing Machinery's Committee on Professional Ethics.
- Angwin, J., Larson, J., Mattu, S., and Kirchner, L. (2016). Machine bias. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. ProPublica.
- Duhigg, C. (2012). How companies learn your secrets. <https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html>. The New York Times Magazine.
- Friedler, S. A., Scheidegger, C., and Venkatasubramanian, S. (2021). The (im)possibility of fairness: Different value systems require different mechanisms for fair decision making. *Communications of the ACM*, 64(4):136 – 143.
- Gigerenzer, G. (2017). Can search engine data predict pancreatic cancer? *BMJ*, 358.
- Sørensen, H. K. and Johansen, M. W. (2021). Invitation til de datalogiske fags videnskabssteori. Lærebog til brug for undervisning ved Institut for Naturfagernes Didaktik, Københavns Universitet. Under udarbejdelseyear.
- The Tech Pledge (2019). The tech pledge. <https://www.techpledge.org/>. Techfestival 150.