

Demokratiske algoritmer

2022-02-17

Contents

1	Om rapporten	5
1.1	Kort sammendrag	5
1.2	Forfattere	5
1.3	Referansegruppe	6
1.4	Data	6
1.5	Finansiering	6
2	Bakgrunn og motivasjon	7
2.1	Byråkratisk omveltning	7
2.2	Rettferdighet og legitimitet	8
3	Oppfatninger om NAV	11
3.1	Erfaring med og kjennskap til NAV	11
3.2	Tillit	12
3.3	Byråkratisk kompetanse	13
3.4	Tiltro til likebehandling i NAV	16
3.5	Oppsummering	18
4	Kunstig intelligens i forvaltningen	21
4.1	Norske innbyggers kjennskap og følelser knyttet til maskinlæring og kunstig intelligens	21
4.2	Bruke kunstig intelligens?	23
4.3	Oppsummering	25
5	Statistisk paritet	27

6	Inputvariabler	31
7	Representativt byråkrati	35

Chapter 1

Om rapporten

1.1 Kort sammendrag

Den pågående automatiseringen av beslutningsprosesser i offentlig forvaltning representerer en omveltning innenfor byråkratisk myndighetsutøvelse. Tilgang på store mengder relevant digital data og økende muligheter for å behandle informasjonen gjør at oppgaver som tidligere måtte behandles manuelt kan overlates til hel- eller halvautomatiserte prosesser med vesentlig redusert menneskelig inngripen. På den ene siden gir denne utviklingen store effektiviseringsmuligheter og potensial for offentlige besparelser. På den andre siden er ivaretagelsen av forvaltningens legitimitet i befolkningen et risikoaspekt i denne utviklingen. Det overordnede målet med denne rapporten er å studere ut fra et demokratiperspektiv om, og i så fall hvordan, oppfattelsen av NAV som institusjon blant innbyggere i Norge påvirkes av en overgang til økt grad av automatisert saksbehandling.

1.2 Forfattere

Sveinung Arnesen er Forsker I og faglig leder for Demokrati og innovasjon ved NORCE, og førsteamanuensis II ved Institutt for administrasjons- og organisasjonsvitenskap, UiB. PhD-graden ble avlagt ved Institutt for sammenliknende politikk, UiB. Arnesen er Norges nasjonale koordinator for Den europeiske samfunnsundersøkelsen (ESS). ORCID. Google Scholar.

Mikael P. Johannesson er forsker ved NORCE, og PhD-kandidat ved Institutt for sammenliknende politikk, UiB. Han har bred erfaring med eksperimentelle metoder, maskinlæring (inkludert deep learning), og surveyforskning. Johannesson har utviklerkompetanse i statistikkprogrammet R, samt erfaring med Python (inkludert TensorFlow og Keras).

1.3 Referansegruppe

Anne Lise Fimreite er professor ved Institutt for administrasjons- og organisasjonsvitenskap, UiB. Hun har tidligere ledet den forskningsrådsfinansierte evaluering av NAV-reformen og har arbeidet mye med styringsutfordringer i flernivåsystem. Hun har også nylig vært medlem av det offentlige utvalget som i 2019 leverte forslag til ny forvaltningslov (NOU 2019:5) og har egen erfaring fra offentlig forvaltning som prorektor ved UiB i fire år fra 2013 til 2017.

Jacob Aars er professor ved Institutt for administrasjons- og organisasjonsvitenskap, UiB, og har ledet den NFR-finansierte evalueringen av NAV-reformen (tok over da Fimreite gikk inn i rektoratet ved UiB). Han har blant annet forsket på lokaldemokrati og tilfredshet med offentlige tjenester.

1.4 Data

Datagrunnlaget for denne rapporten er samlet inn i Norsk medborgerpanels runde 22 og 23. Norsk medborgerpanel er en internettbasert undersøkelse om nordmenns holdninger til viktige samfunnstema. Panelet drives av samfunnsforskere ved Universitetet i Bergen og NORCE, og er et non-profit prosjekt utelukkende benyttet til forskningsformål. Deltakerne representerer et tverrsnitt av den norske befolkningen, som noen ganger i året inviteres til å si sin mening i viktige spørsmål om norsk samfunn og politikk.

Panelet blir driftet av Digital samfunnsvitenskapelig kjernefasilitet (DIGSS-CORE).

Samlet antall respondenter som svarte på deler av denne undersøkelsen var 3971. Typisk antall respondenter for hvert spørsmål er om lag 2000.

Metoderapport for data kan lastes ned her. Data er åpent tilgjengelig for forskere, og kan lastes ned ved å kontakte Sikt – Kunnskapssektorens tjenesteleverandør (tidligere kjent som NSD).

1.5 Finansiering

Forskningsrapporten er finansiert av NAV Forskning og Utvikling. Finansieringen er bidragsfinansiert, det vil si uten krav til motytelse.

Chapter 2

Bakgrunn og motivasjon

2.1 Byråkratisk omveltning

Den pågående automatiseringen av beslutningsprosesser i offentlig forvaltning representerer en omveltning innenfor byråkratisk myndighetsutøvelse. Tilgang på store mengder relevant digital data og økende muligheter for å behandle informasjonen gjør at oppgaver som tidligere måtte behandles manuelt kan overlates til hel- eller halvautomatiserte prosesser med vesentlig redusert menneskelig inngripen (Zarsky, 2016). På den ene siden gir denne utviklingen store effektiviseringsmuligheter og potensial for offentlige besparelser (?). Den representerer også en mulighet for å utvikle bedre, evidensbaserte beslutninger, som i sin tur kan bidra til å bevare tilliten og legitimiteten til offentlig forvaltning. På den andre siden er nettopp ivaretagelsen av forvaltningens legitimitet i befolkningen også et risikoaspekt i denne utviklingen. En frykt er at feil bruk kan lede til utfall som negativt forskjellsbehandler svakerestilte grupper i samfunnet, som igjen underminerer systemtilliten.

En arbeidsgruppe oppnevnt av den tidligere amerikanske president Barack Obama publiserte rapporter hvor de uttrykte bekymring for “kode-diskriminering» i automatiserte beslutninger, hvor diskriminering av sosiale grupper oppsto som en utilsiktet følge av måten stordatateknologi er strukturert og brukes. Dystopiske skildringer av “svart boks”-samfunn maler et skremmende bilde av et framtidssamfunn der innbyggernes skjebner blir bestemt av skjulte, upresise, og diskriminerende automatiske beslutningsprosesser (Barocas and Selbst, 2016; Pasquale, 2015). I de tilfeller oppmerksomheten når ut til allmennheten, har fokus tendert å handle om hvordan beslutningene slår ulikt ut sosiale grupper. Det amerikanske nyhetsmagasinet ProPublica viste hvordan prediksjonsmodeller som brukes til å forutsi gjentakelsesfare for lovbrudd blant fengselsinnsatte systematisk kategoriserte svarte insatte oftere enn hvite feilaktig som personer med høy risiko for å begå en ny forbrytelse når

de løslates fra fengselet (Angwin et al. 2016).

Opplevd diskriminering fra myndighetenes side mot sosiale grupperinger er ikke noe nytt, og spesielt ikke i USA hvor automatiserte beslutninger har fått mest oppmerksomhet til nå. I Norge har vi mindre forskjeller mellom folk, både økonomisk, politisk og sosialt. Norge har også en høyt kompetent og effektiv offentlig forvaltning som jevnt over nyter høy tillit i befolkningen. I overgangen til økt automatisering i forvaltningen er det viktig at tilliten og legitimiteten til offentlig forvaltning opprettholdes.

NAV er ledende i utviklingen av digitale tjenester og verktøy (Hansen, Lundberg, and Syltevik 2018), og utvikler systemer som kan nyttiggjøre seg framskritt som gjøres innenfor databehandling og analyse. Beslutningsprosesser som benytter seg av kunstig intelligens vil være en del av løsningen for at NAV skal oppnå samfunnsoppdraget sitt om å bidra til at flere kommer i arbeid og færre på stønad, og samtidig sørge for at de som trenger det, får rett ytelse til rett tid gjennom en pålitelig og effektiv forvaltning. Kunstig intelligens kan brukes både i automatiserte beslutningsprosesser, og som beslutningsstøtte for saksbehandlere. Et sentralt kjennetegn ved kunstig intelligens er at slike systemer etterligner, erstatter og utvider menneskelig intelligent handling, og menneskelig beslutningstaking og vurdering (Mikkelsen 2019). Mulige områder hvor kunstig intelligens kan benyttes i NAV er blant annet for å beregne sannsynlighet for at den arbeidsledige trenger bistand fra NAV; til å bestemme om en person i sykefravær skal kalles inn til oppfølgingsmøte fra NAV; og til å anbefale arbeidsrettede tiltak. På veien mot bedre tjenester er det viktig at man har med seg brukerne – det vil si innbyggerne i Norge – og lager ansvarlige systemer som gir lik og rettferdig behandling uavhengig av sosial status.

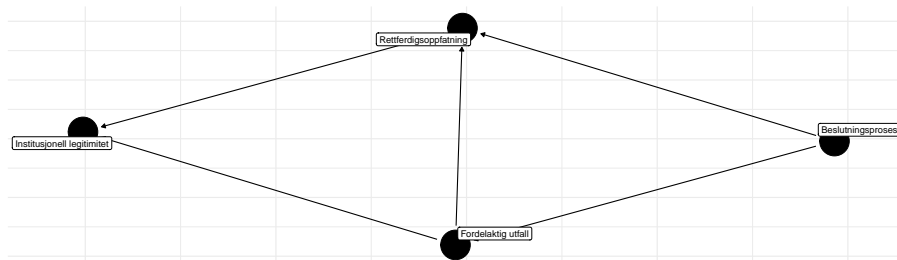
2.2 Rettferdighet og legitimitet

En massiv litteratur på prosedyrerettferdighet med utspring fra sosialpsykologi (Tyler and Lind 1992) har hatt stor innvirkning på hvordan vi forstår relasjonene mellom innbyggere og myndighetene, og hvordan myndighetene bør forholde seg i møte med innbyggerne. Når beslutningsprosessene dreier i retning av mer automatisering, fungerer denne litteraturen som et velegnet rammeverk for å undersøke empirisk om, og i så fall hvordan, relasjonene mellom innbyggerne og myndighetene vil påvirkes av denne utviklingen. Det vi vet fra eksperimentell forskning på politisk atferd er at både aspekter ved prosessen og utfallet i seg selv påvirker rettferdighetsoppfatningen av beslutningen og i sin tur villigheten til å akseptere beslutningen (se figur under). På kort sikt handler det om å skaffe aksept for enkeltbeslutninger. I et mer overordnet perspektiv dreier det seg om systemstøtte; om å sikre tillit og legitimitet til styresmaktene, og opprettholde tilfredsheten med demokratiet som styresett. Legitimitet forstår vi her som makten til å få noen til villig å føye seg etter en beslutning (Weber 2009), som i sin tur gir myndighetene den autoriteten de trenger for å styre effektivt

uten bruk av sanksjoner (Tyler 2006). Demokratisk legitimitet viser til den legitimiteten som vinnes ved at beslutningene utgår fra folkeviljen (Rosanvallon 2011). Både tillit og legitimitet omhandler relasjonen mellom innbyggere og myndighetene, og faller under det bredere konseptet om politisk støtte (Easton 1965).

Prosessrelaterte spørsmål som har blitt studert er blant annet om rettferdighet-soppfatningen og aksepten av beslutningen påvirkes av forhold som er sentrale i demokratiske systemer. Slike forhold kan for eksempel være grad av åpenhet rundt beslutningsprosessen (De Fine Licht et al. 2014), mulighet for direkte påvirkning på en avgjørelse (Esaiaasson, Gilljam, and Persson 2012; Arnesen 2017; Christensen, Himmelroos, and Setälä 2020), hvem beslutningstakerne er, og hvor godt disse beslutningstakerne gjenspeiler befolkningen med tanke på sosial bakgrunn, også kjent som deskriptiv representasjon (Arnesen and Peters 2018; Clayton, O'Brien, and Piscopo 2019). Videre er et gjennomgående funn at dersom utfallet går imot ens egne ønsker, blir prosessen diskreditert (Esaiaasson et al. 2019; Marien and Kern 2018; Arnesen, Broderstad, et al. 2019; Arnesen, Bergh, et al. 2019).

Det er viktig å studere rettferdighet fra et statsvitenskapelig perspektiv fordi oppfatninger av rettferdighet antas å påvirke institusjonell legitimitet (Tyler 2003). Dette begrenser seg ikke bare til input-siden av det politiske systemet hvor politikk vedtas, men også output-siden i forvaltningen hvor politikk settes ut i live (Krislov 2012; Rosanvallon 2011; Rothstein 2011).



Forskningen på kunstig intelligens og rettferdighet er raskt voksende. Flere definisjoner brukes om rettferdighet, og de fleste av dem er basert på forhold mellom sanne/falske positive og sanne/falske negative (Verma and Rubin 2018). Alexandra Chouldechova Chouldechova et al. (2018) viser teoretisk og empirisk hvordan to velbrukte definisjoner av rettferdighet umulig kan oppnås samtidig i visse tilfeller. Hvilke definisjoner bør prioriteres når man står overfor slike avveininger? Chouldechovas studie viser med tydelighet at det ikke er gjort i en håndvending å lage rettferdige prediksjonsmodeller. Tvert imot er det komplisert utfordring som krever oppmerksomhet om konkret kontekst. Vi vet fortsatt lite om hvilke definisjoner som resonnerer blant innbyggerne, og i hvilken grad de modereres av kontekst eller innbyggernes sosiale bakgrunn eller politiske holdninger. Vi vet fra samfunnsforskning at hva som oppfattes som rettferdig kan variere med sosial identitet og kultur, politiske holdninger, og personlige

karaktertrekk, og det er derfor viktig å gjøre konkrete empiriske studier på realistiske problemstillinger før slike prediksjonsmodeller tas i bruk. I [kapittel] @ref(#paritet) tar vi for oss et realistisk scenario for NAV hvor vi setter opp to motstridende rettferdighetshensyn knyttet til hvem som skal få tilbud om dialogmøte.

På tross av et økende krav om at innbyggerne må involveres og konsulteres (Balaram, Greenham, and Leonard 2018), har demokrati- og opinionsforskere i liten grad studert hvordan overgangen til økt bruk av kunstig intelligens i forvaltningen kan påvirke tillit og legitimitet. Viktige unntak er De Fine Licht & De Fine Licht (2020) som studerer rollen som åpenhet når det gjelder hvordan allmennheten oppfatter kunstig intelligens-beslutninger som legitim, og Binns et al (2018) som foretar eksperimentelle studier som undersøker folks oppfatning av rettferdighet i algoritmiske beslutninger. Vår kunnskap om hvordan overgangen vil påvirke innbyggernes oppfatninger av forvaltningen er imidlertid fortsatt svært begrenset, og behovet for befolkningsrepresentative studier med et demokratiperspektiv er stort. Det er motivasjonen for å lage denne rapporten.

Rapporten presenterer resultatene fra en spørreundersøkelse gjennomført på et befolkningsrepresentativt utvalg av innbyggere i Norge. Undersøkelsen tar for seg generelle holdninger til og kunnskap om maskinlæring og kunstig intelligens i befolkningen. Bredere spørsmål knyttet til relasjonen mellom NAV og innbyggerne – uavhengig av tematikken om maskinlæring og kunstig intelligens – blir også analysert for å kontekstualisere de mer spesifikke spørsmålene og eksperimentene knyttet til bruk av maskinlæring og kunstig intelligens i NAV. Deretter fokuserer vi på holdninger som har relevans for en framtid hvor maskinlæring og kunstig intelligens har en sentral rolle i. Dette gjelder problemstillinger knyttet til konkrete, aktuelle situasjoner i NAV, såvel som mer overordnede spørsmål om maskinlæring og kunstig intelligens i forvaltningen.

Chapter 3

Oppfatninger om NAV

I dette kapitlet ser vi nærmere på relasjonen mellom NAV og innbyggerne i Norge. Kapitlet danner grunnlaget for å forstå konteksten når vi senere fokuserer mer spesifikt på maskinlæring og kunstig intelligens i organisasjonene. Vi måler folks tillit til NAV, samt deres erfaring med og kjennskap til NAV. Vi måler også i hvilken grad innbyggerne opplever at de forstår hvordan organisasjonen fungerer, og i hvilken grad de opplever at deres interesser blir ivaretatt i systemet.

3.1 Erfaring med og kjennskap til NAV

Et flertall av innbyggerne i Norge har hatt befatning med NAV i en eller annen form. Noen kontakter er lite personlige, som for eksempel når man mottar barnetrygd. Andre krever mer kontakt med NAV, og gjerne personlig kontakt med en saksbehandler. I vårt utvalg har godt over halvparten minst en gang vært i personlig kontakt med NAV.

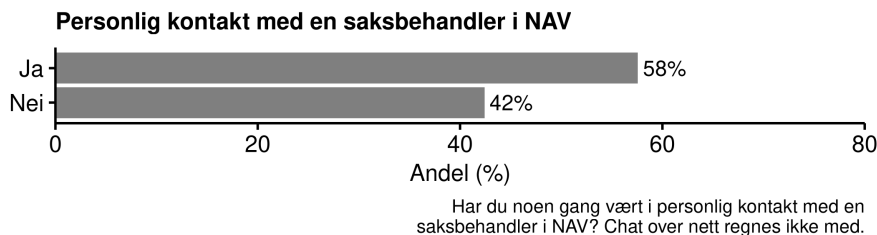
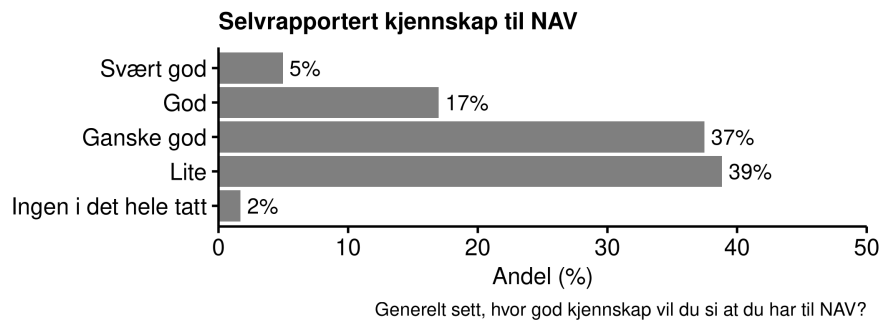


Figure 3.1: Personlig kontakt med NAV

Fire av ti innbyggere i Norge har liten eller ingen kjennskap til NAV, mens seks av ti har ganske god, god, eller svært god kjennskap til institusjonen.



3.2 Tillit

Norge er kjent som et land der myndighetene nyter høy tillit i befolkningen. Dette bekreftes også i vår undersøkelse for NAV sin del. Fire av ti innbyggere i Norge har høy eller svært høy tillit til NAV, mens bare en av seks har liten eller ingen tillit. På en skala fra 1 til 5 der 1 er lavest og 5 er høyest, er gjennomsnittsverdien på **xx**.

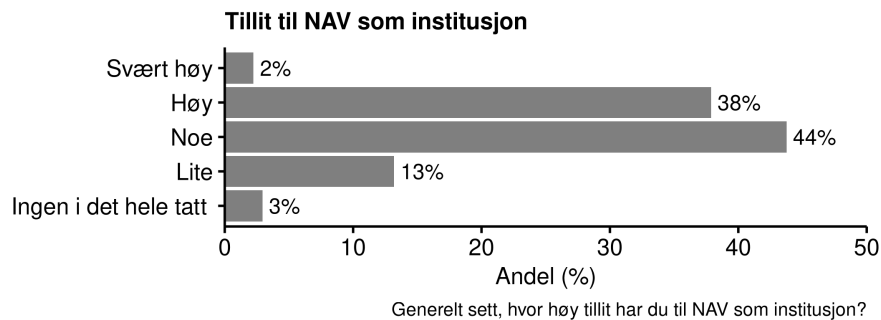


Figure 3.2: Tillit til NAV

Tillit er noe som tar lang tid å bygge opp, men som fort kan rives ned. Det er et godt utgangspunkt at NAV nyter tillit i befolkningen, og ikke desto viktigere at denne relasjonen ivaretaes parallelt med at organisasjonen endrer seg og utvikler morgendagens forvaltning.

1. SAMMENLIKNE MED TILLIT TIL ANDRE ORGANISASJONER
2. TILLIT BRUTT NED PÅ ERFARING MED OG KJENNSKAP TIL NAV

3.3 Byråkratisk kompetanse

I statsvitenskapen opererer man med et begrep som på engelsk heter *political efficacy*. Vi kan gjerne oversette begrepet på norsk til *politisk kompetanse*. Politisk kompetanse viser til en persons selvopplevde evne til å forstå politikk (*internal political efficacy*), og personens opplevelse av å kunne påvirke politiske prosesser (*external political efficacy*). Vi ser at intern og ekstern politisk kompetanse henger sammen med politisk deltakelse, med tilfredshet med demokratiet, tillit til institusjoner, blant annet.

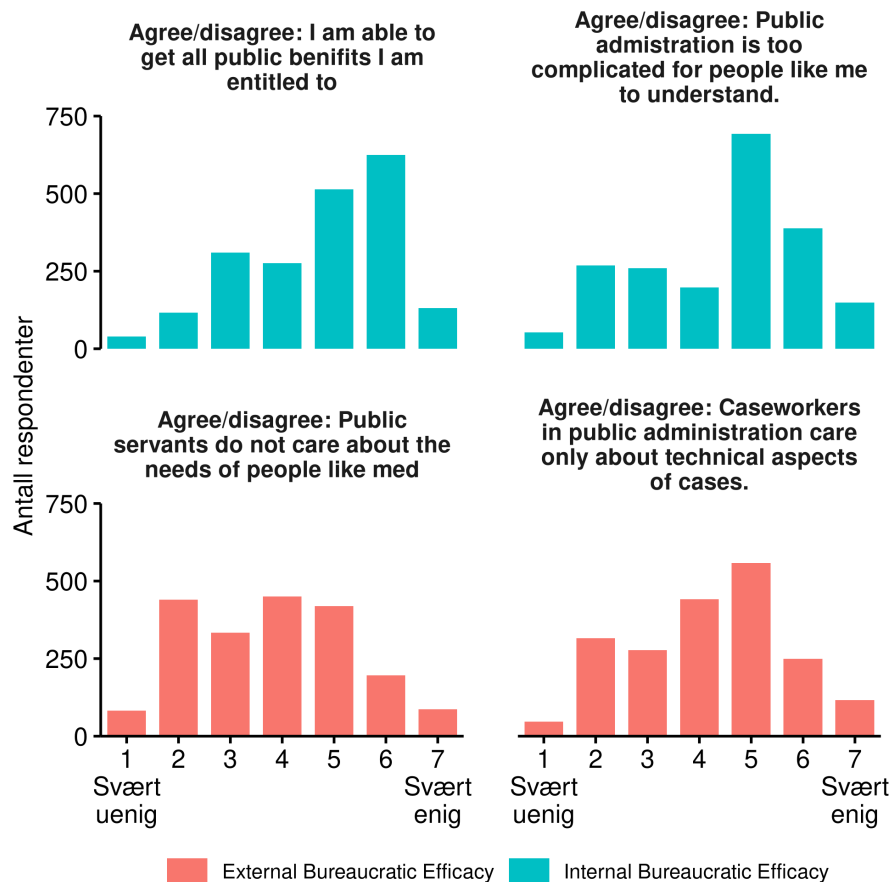
På samme måte som at innbyggerne har ulike evner til å forstå politikk og påvirke politiske prosesser, har de også ulike evner til å forstå forvaltningen. I motsetning til i politisk arbeid er det ikke et mål at innbyggerne nødvendigvis skal kunne påvirke en byråkratisk prosess, men det er likefullt en kjennsgjerning at noen personer er flinkere til å følge opp saker på vegne av seg selv eller pårørende, og slik sett er bedre i stand til å ivareta sine interesser i saker som angår dem.

For å avdekke hvordan disse ferdighetene fordeler seg i befolkningen har vi spurt respondentene om deres byråkratiske kompetanse - intern og ekstern. Intern byråkratisk kompetanse måler vi ved hjelp av to påstander som de skal si seg enig eller uenig i:

1. Jeg er i stand til å skaffe alle offentlige ytelser, tjenester, og tillatelser som jeg har rett på.
2. Den offentlige forvaltningen er så innviklet at folk som meg ikke kan forstå hva som foregår innad i ulike etater, direktorat, kommuner, og så videre.

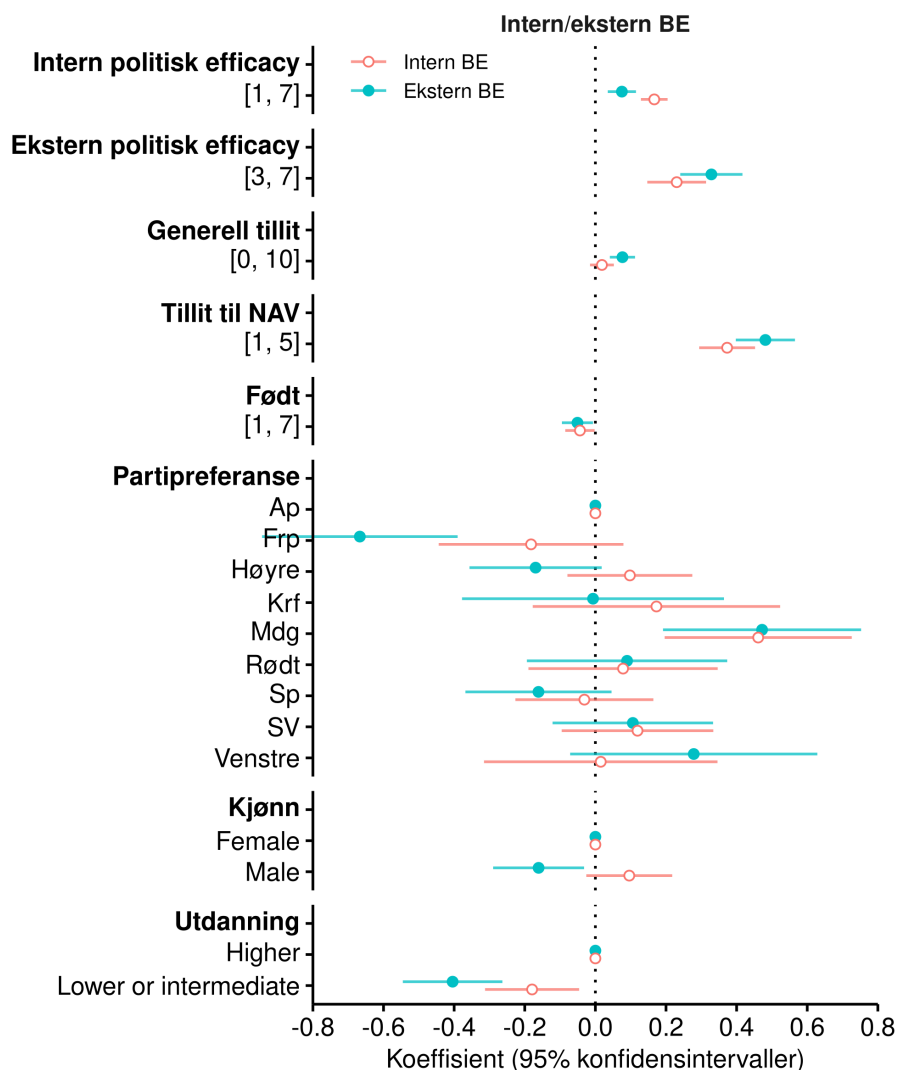
Ekstern byråkratisk kompetanse måler vi ved hjelp av to nye påstander:

1. De som jobber i den offentlige forvaltningen bryr seg ikke om hvilke behov folk som meg har.
2. Saksbehandlere i den offentlige forvaltningen er bare interessert i tekniske aspekter ved saken, ikke hva de det berører faktisk ønsker.



Figurene viser at det er ganske stor spredning i svarene. Flertallet av innbyggerne er enige i påstanden at de får alle ytelser som de har rett på. Et flertall svarer samtidig at de opplever byråkratiet som vanskelig å forstå. Flertallet er uenige i at de som jobber i offentlig forvaltning ikke bryr seg om folks behov, men det er også et stort mindretall som er enige i denne påstanden. Et lite flertall mener at saksbehandlere bare er interessert i tekniske aspekter ved saken.

Dette er befolkningen i sin helhet. Når vi bryter svarene ned på sosiopolitiske undergrupper ser vi at dem som har lav byråkratisk kompetanse også har lav politisk kompetanse (efficacy). Det vil si at de som har tillit til NAV er de som opplever at de får ytelsene de har rett på, som føler at de forstår de byråkratiske prosessene, som tenker at de som jobber i offentlig sektor bryr seg om folks behov, og som ikke bare forholder seg til tekniske aspekter i saksbehandlingen.

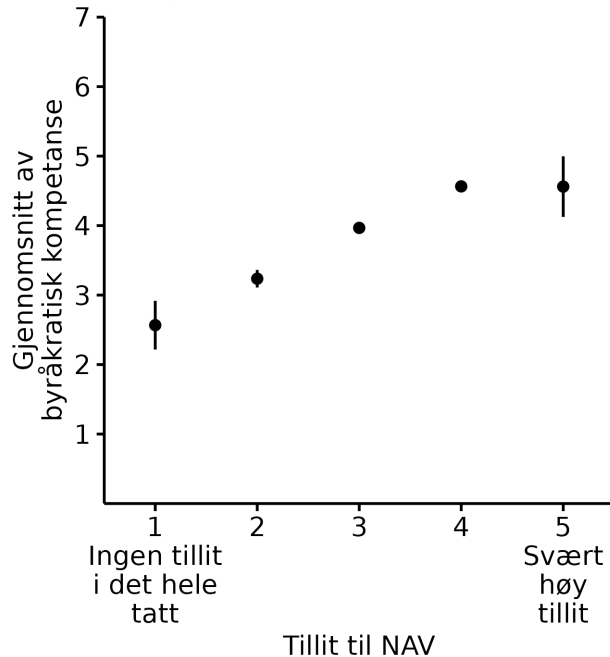


Byråkratisk kompetanse varierer også til dels mye mellom folk når vi deler dem inn i hvilket parti de ville ha stemt på dersom det var stortingsvalg i morgen. Figuren under viser at det er størst forskjell på de som stemmer Fremskrittspartiet og de som stemmer Miljøpartiet de grønne, spesielt når det gjelder ekstern byråkratisk kompetanse.

Menn scorer noe lavere enn kvinner på ekstern byråkratisk kompetanse, og det samme gjør personer uten høyere utdanning sammenliknet med personer med høyere utdanning. De scorer også noe lavere på intern byråkratisk kompetanse, men her er forskjellen mindre (men fortsatt statistisk signifikant).

Tillit til NAV samvarierer også sterkt med byråkratisk kompetanse. Figuren

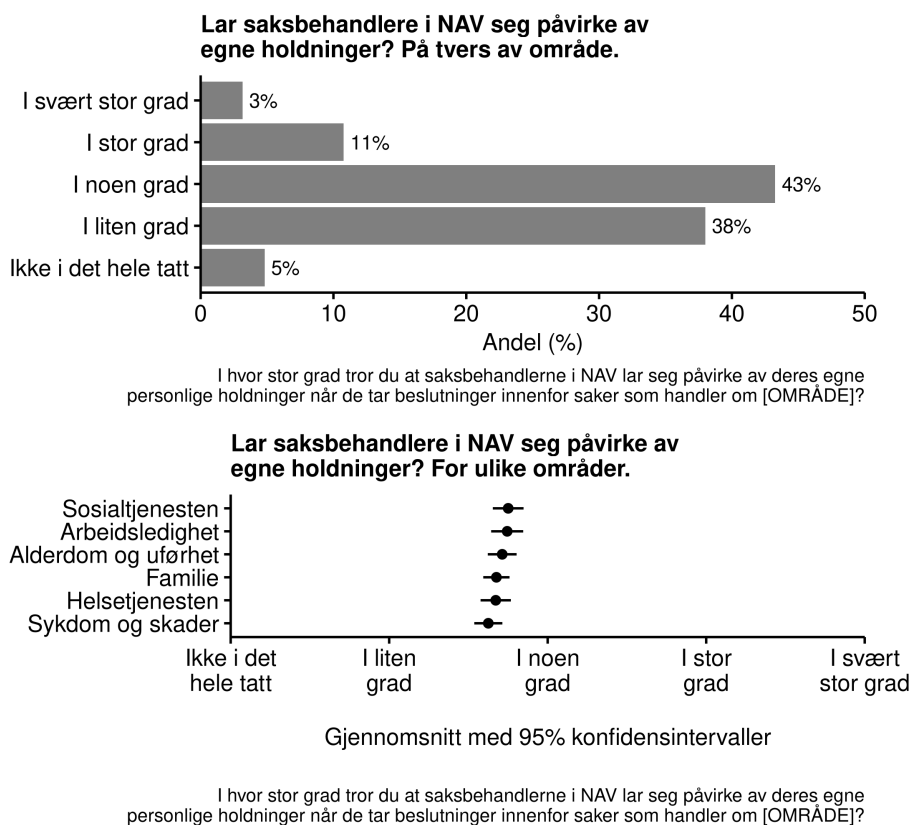
under viser mer dette mer detaljert. Vi observerer en lineær sammenheng mellom de to variablene, som flater ut først blant de som har svært høy tillit til



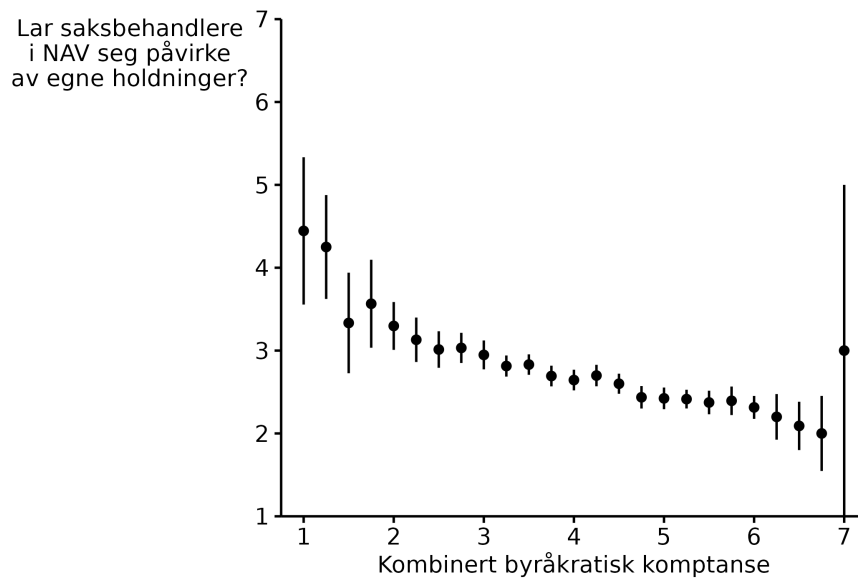
NAV.

3.4 Tiltro til likebehandling i NAV

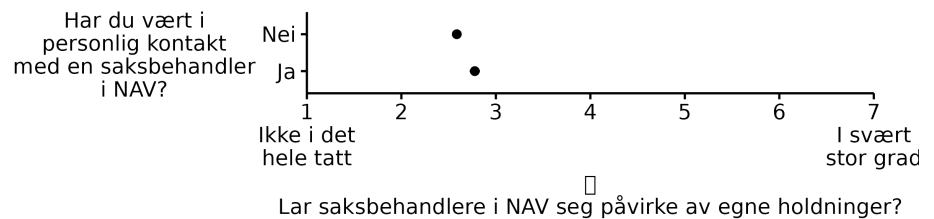
Et styrende prinsipp i norsk forvaltning er nøytralitet: Saksbehandlere skal utøve sitt mandat uten å la egne holdninger komme i veien og påvirke beslutningsprosessen. Slik uhildet behandling av innbyggerne står også sentralt i forskning på hvilke egenskaper ved byråkratiet som underbygger legitimitet og rettferdighetsoppfatninger. Når vi spør innbyggerne i Norge hvordan de oppfatter at saksbehandlerne i NAV lar seg påvirke av egne holdninger, finner vi at de tror det forekommer i noen grad. Det er liten forskjell på de ulike områdene innenfor NAV.



Figuren under viser også at byråkratisk kompetanse (omtalt i forrige kapittel) samvarierer med oppfatningen om at saksbehandlere lar seg påvirke av egne holdninger: Jo høyere byråkratisk kompetanse en innbygger har, desto mindre mener man at saksbehandlerne lar seg påvirke av egne holdninger.



I utvalget mener de som har vært i personlig kontakt med en saksbehandler i NAV i litt større grad at saksbehandlerne lar seg påvirke av egne holdninger. Forskjellen er imidlertid ikke stor, og **HELLER IKKE STATISTISK SIG-**



NIFIKANT.

3.5 Oppsummering

Dette kapitlet har gitt oss oppdaterte data på innbyggere i Norges relasjon til NAV. Vi ser at

- over halvparten av innbyggerne har vært i personlig kontakt med saksbehandler i NAV, men mange oppgir likevel at de har liten kjennskap til NAV.
- innbyggerne har middels til høy tillit til NAV. Det er få som oppgir å ha svært høy tillit eller ingen tillit i det hele tatt.
- flertallet av innbyggerne opplever at de er i stand til å få de tjenestene de har krav på fra det offentlige. Dette på tross av at mange opplever forvaltningen som krevende å forstå.

- innbyggerne i liten til noen grad mener at saksbehandlere i NAV lar seg påvirke av egne holdninger. Det skiller lite mellom hvilke områder av NAVs ansvarsområde man spør etter.

Chapter 4

Kunstig intelligens i forvaltningen

4.1 Norske innbyggers kjennskap og følelser knyttet til maskinlæring og kunstig intelli- gens

Maskinlæring som databehandlingsmetode er relativt fersk. Med økt datakraft og økt tilgang på data, har bruk av maskinlæring bredt om seg innenfor datavitenskapelige miljøer.

I befolkningen for øvrig er det antatt lite kjennskap til maskinlæring og kunstig intelligens. I forkant av spørsmålene får derfor alle en liten introduksjon til hva det er.

De blir opplyst at å bruke maskinlæring innebærer å få datamaskiner til å lære seg å løse oppgaver basert på et datamateriale. Ofte kan datamaskinen bli ekstremt treffsikker, men det krever typisk veldig mye datamateriale. Maskinlæring er i dag grunnlaget for alt fra automatisk stemmegjenkjenning til førerløse biler. Den offentlige forvaltningen, inkludert NAV, bruker i enkelte tilfeller maskinlæring for å hjelpe med å ta beslutninger i saker de har ansvar for. Formålet er å redusere kostnader og behandlingstid, og å gjøre beslutninger bedre og mer treffsikre. Et eksempel kan være å lære en datamaskin å forutsi omtrent hvor lenge en person vil være sykmeldt, basert på informasjon om sykdommen og personen. Det kan en saksbehandler da bruke for å velge passende tiltak.

Etter denne korte introduksjonen kartlegger vi deres kjennskap og holdninger rundt bruk av maskinlæring i den offentlige forvaltningen.

Bare en av syv innbyggere oppgir at de har god eller svært god kjennskap til

maskinl ring og kunstig intelligens, mens nesten to tredjedeler sier at de har liten eller ingen kjennskap til det i det hele tatt.

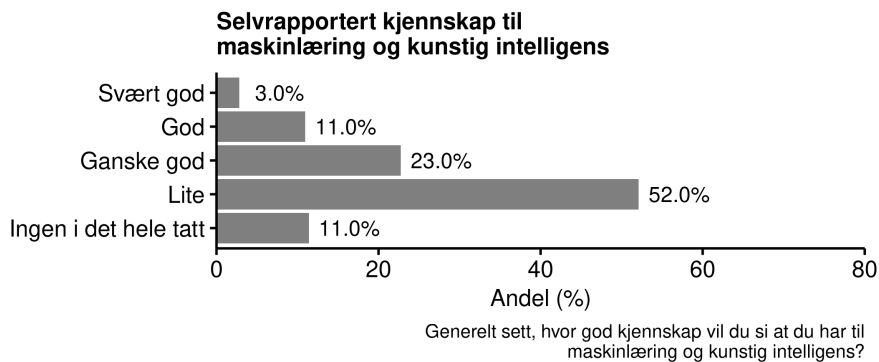


Figure 4.1: Kjennskap til maskinl ring

4.1.1 Bekymring for maskinl ring

Innbyggerne er delt i synet p  grad av bekymring knyttet til bruk av maskinl ring og kunstig intelligens i den offentlige forvaltningen.

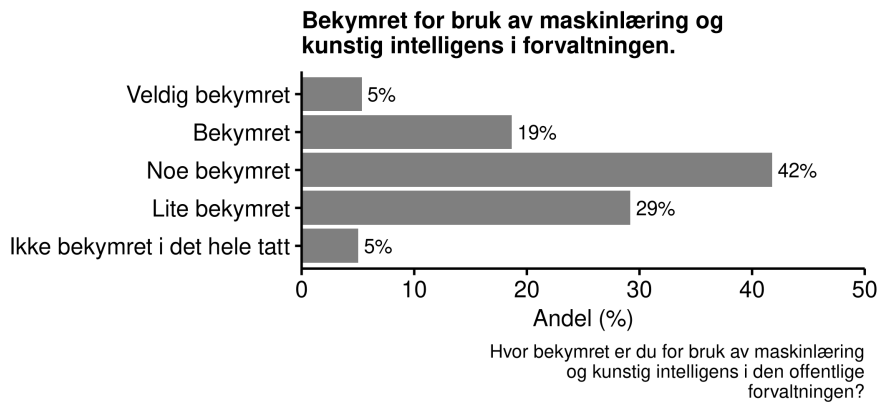


Figure 4.2: Bekymring for maskinl ring

1.  PENT TEKSTSVAR MED BEGRUNNELSE FOR SVAR OM BEKYM-RING
2. ER DET MANGEL P  KUNNSKAP SOM LEDER TIL BEKYM-RING?

4.1.2 Interesser ivaretatt med maskinlæring?

Det er naturlig å se maskinlæring og kunstig intelligens i NAV i sammenheng med spørsmål som ligger nær opptil byråkratisk kompetanse. Vil folk oppleve at det er lettere eller vanskeligere å forstå hvordan byråkratiet fungerer? Vil deres interesser ivaretas bedre eller dårligere når maskinlæring brukes i NAV?

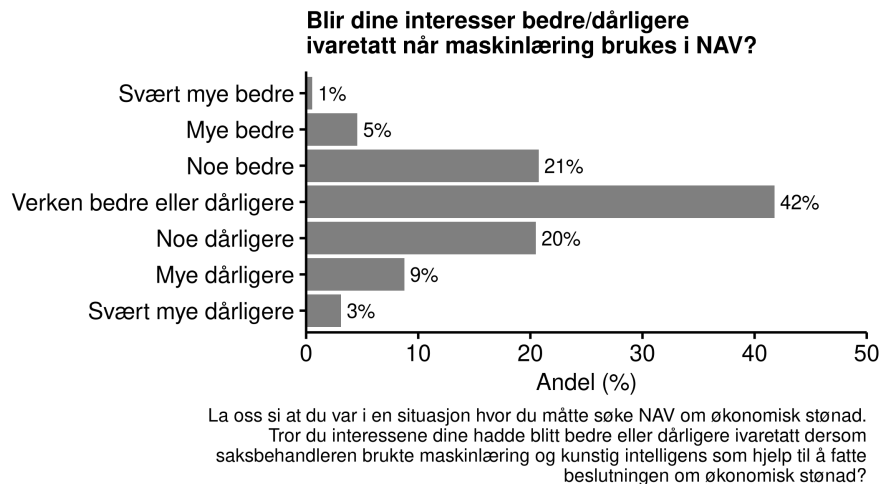


Figure 4.3: Maskinlæring i NAV

Det mest vanlige svaret var midtkategorien ‘verken bedre eller dårligere’. For øvrig fordelte svarene seg normalt rundt denne midtkategorien. I spørreundersøkelser kan midtkategorier i slike bipolare skalaer (les: bedre vs. dårligere) ofte skjule at respondentene ikke har noen mening om spørsmålet.

1. VI DELER DERFOR OPP SVARENE SLIK AT VI KAN UNDERSØKE OM SVARENE TIL RESPONDENTENE VARIERER ETTER HVOR GOD KJENNSKAP DE HAR TIL MASKINLÆRING.

4.2 Bruke kunstig intelligens?

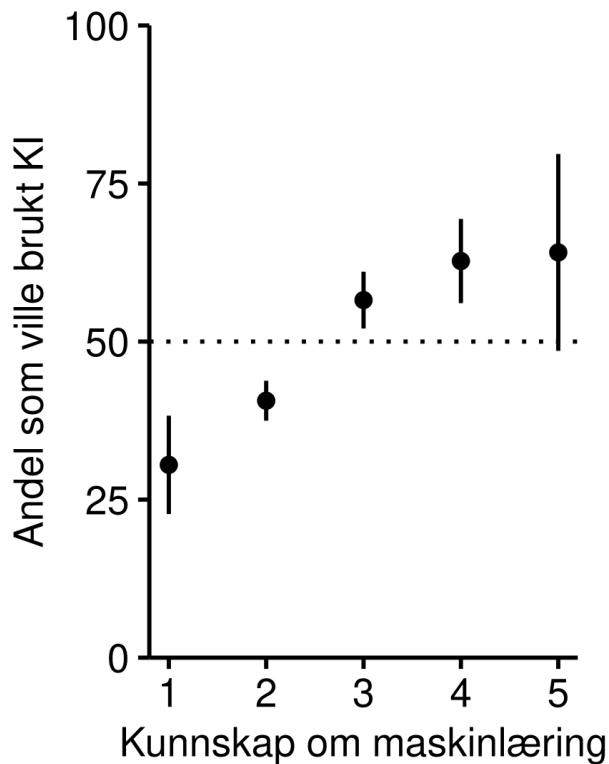
I mange beslutninger i forvaltningen må det utvises skjønn basert på en samlet vurdering av den enkelte saken. Om man tar i bruk kunstig intelligens, ved hjelp av maskinlæring, vil beslutningene antakelig bli mer treffsikker, og dermed øke andelen riktige beslutninger. Samtidig kan heller ikke en datamaskin være helt treffsikker. Det er også grunn til å tro at den gjenværende andelen uriktige beslutninger går mer systematisk ut over noen grupper i samfunnet når

man bruker maskinlæring og kunstig intelligens. Dette fordi det er stor variasjon mellom hvordan menneskelige saksbehandlere utviser skjønn, mens for en datamaskin er det ingen variasjon.

Med dette som bakgrunn spurte vi respondentene hva foretrekker i slike situasjoner: Enten

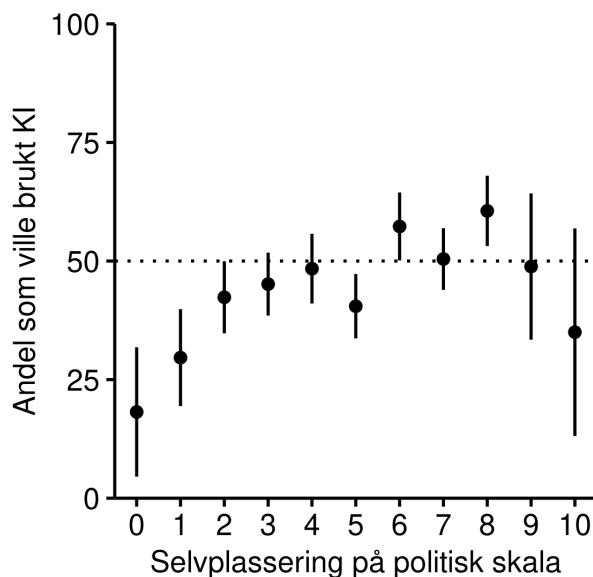
1. Bruke kunstig intelligens, som fører til mange flere riktige beslutninger i bytte mot at det alltid er de samme som blir gjenstand for uriktige avgjørelser, eller
2. ikke bruke kunstig intelligens, som fører til mange færre riktige beslutninger i bytte mot at det varierer hvem som blir gjenstand for uriktige avgjørelser.

Respondentene delte seg på midten i dette spørsmålet, hvor XX PROSENT FORETRAKK Å BRUKE KUNSTIG INTELLIGENS, MENS XX FORETRAKK Å IKKE BRUKE KUNSTIG INTELLIGENS. Figuren under viser at de med lav kjennskap til maskinlæring og kunstig intelligens var mest skeptiske. Det kan altså ha sammenheng med skepsis til det ukjente.



Spørsmålet har også en politisk-filosofisk dimensjon over seg. Premisset som legges til grunn for spørsmålet er at man ved å innføre kunstig intelligens påvirker fordelingen av riktige beslutninger. (Man kan selvsagt diskutere om dette er en riktig virkelighetsbeskrivelse av en forvaltning som bruker kunstig intelligens versus en som ikke gjør det. Vi kommer i denne omgang ikke nærmere inn på dette.) Det blir da et spørsmål om fordeling av goder, og om man er villig til å ofre et lite antall individer som systematisk forfordes med uriktige beslutninger, mot at populasjonen som helhet nyter godt av en høyere andel riktige beslutninger.

Ut fra dette perspektivet gir det mening at de som plasserer seg lengst til venstre er minst villige til å bruke kunstig intelligens. Vi noterer oss også at de som plasserer seg lengst til høyre også er mindre villige til å bruke kunstig intelligens når konsekvensene av bruken presenteres slik som den har blitt gjort i dette konkrete tilfellet.



4.3 Oppsummering

- Mer enn seks av ti innbyggerne i Norge har liten eller ingen kjennskap til maskinlæring og kunstig intelligens
- Innbyggerne er delt i oppfatningen om bruken av maskinlæring og kunstig intelligens i forvaltningen er noe å bekymre seg over
- De som oppfatter at de har god kunnskap om maskinlæring er mer positive til bruk av kunstig intelligens i forvaltningen

- Det er en omvendt U-formet sammenheng mellom selv plassering på politisk høyre/venstre-skala og oppslutning om bruk av kunstig intelligens: Innbyggere som plasserer seg mot midten av det politiske spekteret er mer positive enn de som plasserer seg mot en av endene på skalaen.

Det er viktig å ha i mente at de to siste kulepunktene omhandler svar som blir gitt i en spesifikk “framing” av kunstig intelligens; nemlig en situasjon hvor bruken av kunstig intelligens påvirker beslutningene på en måte som gjør dem mer treffsikre, men samtidig mer systematisk feildiagnostiserer.

Når vi fokuserte spesifikt på en endring av hvordan beslutninger fattes ved bruk av kunstig intelligens kontra uten.

Chapter 5

Statistisk paritet

Et realistisk eksempel hvor maskinlæring kan brukes i forvaltningen er når NAV skal bestemme hvilke sykmeldte som skal få tilbud om dialogmøte med NAV. Et dialogmøte er en samtale mellom NAV og den sykmeldte som anses som positivt for den sykmeldtes muligheter for å komme tilbake i arbeid.

I prinsippet har alle rett på et dialogmøte, men i praksis foregår det en siling hvor det gjøres en vurdering av hvem som har mest nytte av et slikt møte. Bruk av maskinlæring og kunstig intelligens kan i dette tilfellet bidra til bedre estimer for hvem som er i fare for å bli langtidssykmeldt, og derfor kan ha større nytte av et dialogmøte. Derfor er NAV i innledende stadier på å utvikle maskinlæringsmodeller som predikerer sannsynlighet for at en sykemeldt fortsatt vil være sykemeldt 12 uker fram i tid.

Når man bestemmer innretningen på en modell må man foreta prioriteringer. En prinsipielt viktig prioritering handler om man skal ta i bruk såkalt statistisk paritet som rettferdighetsprinsipp på utvalgte egenskaper ved individene det gjelder. Kjønnsparitet er ett eksempel, men det kan også handle om statistisk paritet etter alder, etnisitet, geografi, med mer. Et kjent eksempel innenfor litteraturen om rettferdig bruk av kunstig intelligens er studien som viser hvordan afro-amerikanske fengselsinnsatte sjeldnere blir tilbudt prøveløslatelse enn hva andelen deres skulle tilsi. Dette skjer når avgjørelsen om prøveløslatelse baserer seg på prediksjonsmodeller om risikoen for at den innsatte blir tatt påny for en kriminell handling dersom hen slippes fri (Chouldechova 2017). Å anvende paritetsprinsippet her innebærer å sikre at andelen innsatte som tilbys prøveløslatelse samsvarer med andelen innsatte for hver av de etniske gruppene i fengselet. Fordelen med å bruke statistisk paritet etter etnisitet er at prøveløslatelse blir likt fordelt blant de etniske gruppene, og slik sett kan oppleves som rettferdig fordelt. Utfordringen ved å bruke dette prinsippet er at andre egenskaper ved de innsatte – som for eksempel risikovurderinger om tilbakefall til kriminelle handlinger – blir nedprioritert. Er etnisitet i dette tilfellet så viktig at man bør la det gå på bekostning av risikovurderinger knyttet til tilbakefall?

NAV's tilfelle om dialogmøte er mindre dramatisk enn eksempelet om prøveløslatelse. Samtidig er de prinsipielle problemstillingene de samme. Statistisk paritet innebærer i tilfellet om dialogmøte at modellen sikrer at like mange menn og kvinner skal få tilbud om dialogmøte. Denne prioriteringen vil i så fall gå delvis på bekostning av å prioritere treffsikkerhet med tanke på å invitere de som har størst nytte av et slikt møte.

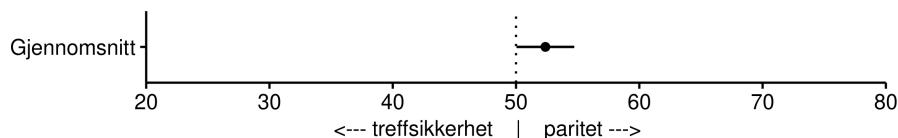
For å studere respondentenes umiddelbare reaksjoner til et slik etisk dilemma knyttet til rettferdig bruk av kunstig intelligens ber vi respondentene se for seg et valg mellom to alternative maskinlæringsmodeller for å velge hvem som skal få tilbud om dialogmøte.

Ingen av modellene er perfekte, men de feiler på ulike måter.

Den første modellen er mest *treffsikker*. Det vil si at det totalt sett er flere sykmeldte med behov for dialogmøte som får tilbudet enn tilfellet er for den andre modellen. Samtidig har modellen en bias til fordel for menn, som gjør at det er flere kvinner med behov for dialogmøte som ikke får tilbudet. Andelen som har behov for dialogmøte *uten å få tilbud* er altså større hos kvinner enn menn.

Den andre modellen sikrer statistisk paritet etter kjønn, nemlig at andelen av de sykmeldte som kalles inn til dialogmøte er like stor henholdsvis for kvinner som for menn. Imidlertid er den mindre treffsikker totalt sett, slik at færre som har behov for dialogmøte blir innkalt. Dette gjelder både kvinner og menn.

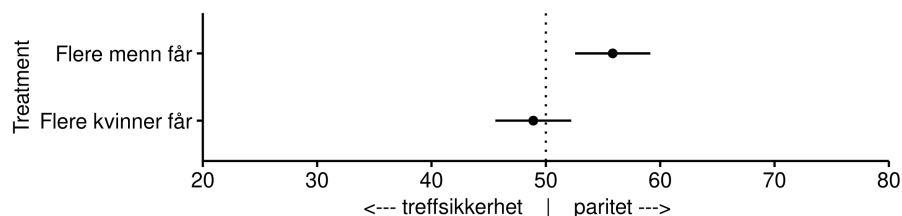
Hvis det står mellom disse to modellene, hvilken modell synes respondentene virker mest rettferdig? Figuren under viser at et knapt flertall foretrekker en modell som vektlegger statistisk paritet. Det vil si at de ønsker å bruke en modell som sikrer likebehandling av kjønn, selv på bekostning av lavere treffsikkerhet totalt sett.



I teksten over står det at den mest treffsikre modellen favoriserte menn. For å undersøke om det har noen innvirkning på svarene hvilket kjønn modellen favoriserer veksler vi på denne beskrivelsen. Halvparten av respondentene får vite at modellen har en bias til fordel for menn, mens den andre halvparten av respondentene får vite at modellen favoriserer kvinner. Spiller det noen rolle hvilket kjønn modellen favoriserer? Resultatene viser at det gjør det.

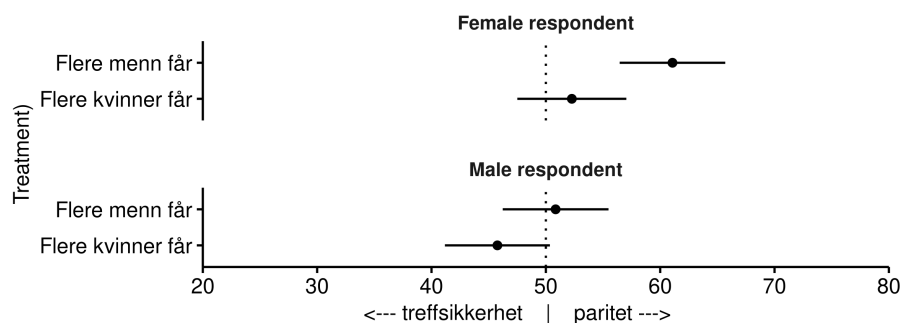
I figuren under ser vi at det er i de tilfeller hvor menn blir fordelaktig behandlet ved bruk av den mest treffsikre modellen at flertallet ønsker å bruke en modell som sikrer likebehandling av kjønn. Det er en signifikant større andel av respondentene som foretrekker paritetsprinsippet når menn har fordel av den treffsikre modellen enn når kvinner har det.

Hva dette skyldes vet vi ikke. Man ser liknende kjønnseffekter i eksperimenter om politisk representasjon, hvor kvinnelige kandidater jevnt over foretrekkes i noe høyere grad enn mannlige kandidater gjør (Schwarz & Coppock 2020). I den litteraturen pekes det på forklaringer om at folk er motivert ut fra et ønske om å kompensere for historisk underrepresentasjon av kvinner i politiske stillinger. Hvorvidt det ligger liknende strukturelle motivasjoner for våre resultater, psykologiske faktorer, eller andre forhold er et interessant forskningsspørsmål som vi ikke har data til å besvare, og som derfor bør studeres videre.

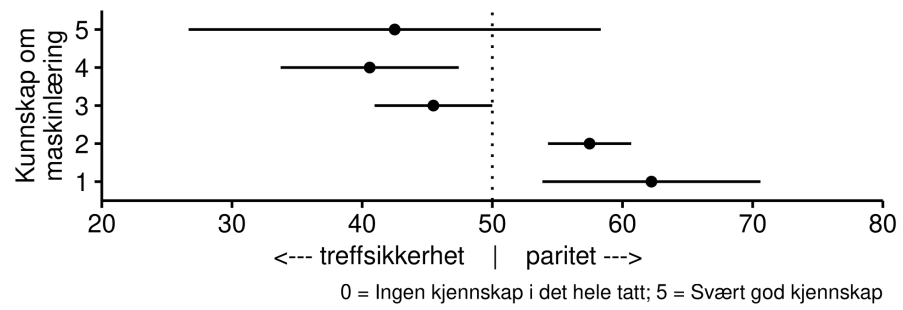


Det vi imidlertid ser, og til forskjell fra litteraturen om politisk representasjon, er at kvinner responderer noe mer på informasjon om hvilket kjønn som kommer best ut av en modell som prioriterer treffsikkerhet. Både menn og kvinner foretrekker paritetsmodellen oftere i de tilfellene kvinnene kommer dårlig ut av treffsikkerhetsmodellen enn i de tilfellene hvor menn kommer dårlig ut av samme modell, men denne effekten er noe sterkere hos kvinner enn menn.

Det er også en generell forskjell blant respondentene i den forstand at kvinner i sterkere grad foretrekker paritetsmodellen enn menn gjør, uavhengig av om det er menn eller kvinner som kommer best ut av det.

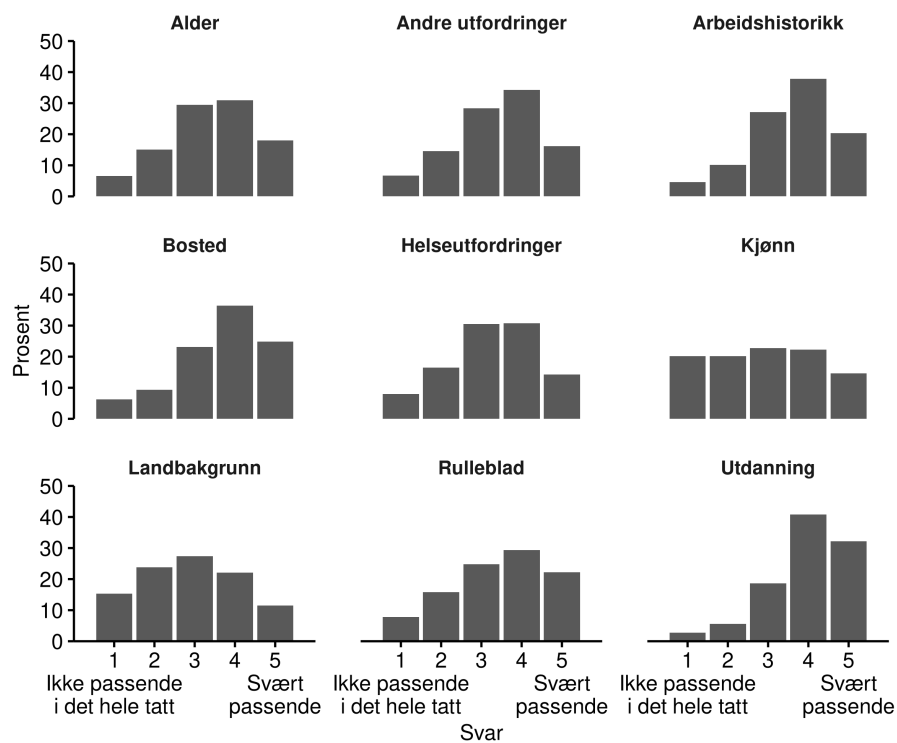
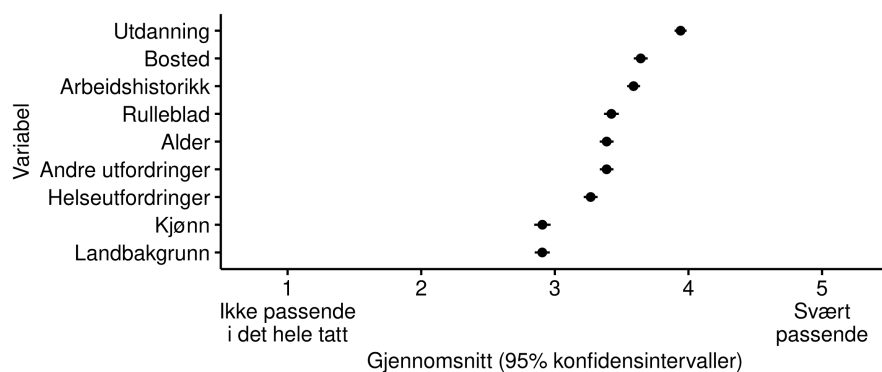


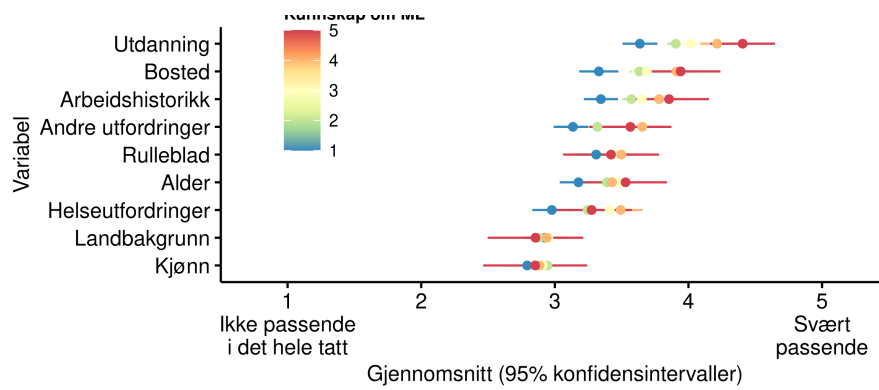
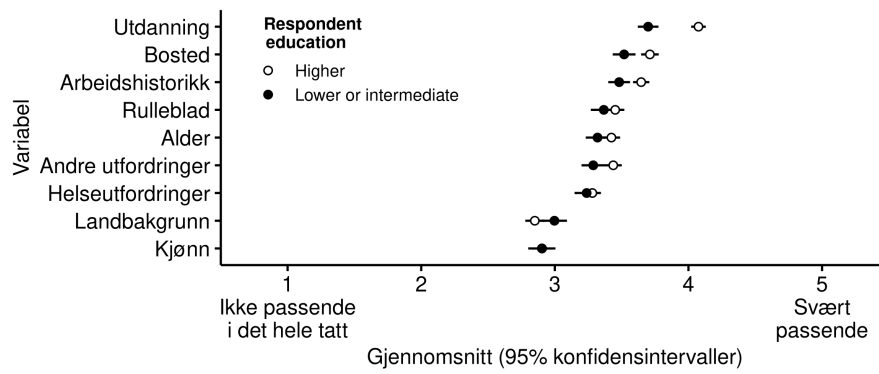
Vi observerer også en modererende effekt av kunnskap om maskinlæring: Jevnt over ser vi at jo høyere kunnskap om maskinlæring, desto høyere andel modellen som prioriterer treffsikkerhet framfor statistisk paritet. Det her er igjen verdt å nevne at slike bivariate, statistiske sammenhenger ikke sier noe om årsakssammenhenger. SKAL VI TA MED DENNE FIGUREN? DEN PASSER IKKE SUPERGODT INN I NARRATIVET, OG ER VANSKELIG Å DISKUTERE.



Chapter 6

Inputvariabler





Chapter 7

Representativt byråkrati

Deskriptiv representasjon er et viktig konsept innenfor studiet av politisk representasjon (Pitkin 1967). Velgere ønsker å bli representert av kandidater som deler deres sosiale bakgrunn, ikke minst fordi de antar at disse kandidatene deler deres politiske interesser og vil ivareta dem på en god måte (Arnesen, Duell, Johannesson 2019).

Mens spørsmål om representasjon er en naturlig del av studier knyttet til input-siden av det politiske systemet (Easton xxxx), er det mindre brukt for å studere output-siden, altså i forvaltningen, hvor politikk skal gjennomføres. Grunnen til dette er tanken om at det er tilstrekkelig med demokratisk innflytelse i utformingen av politikken, så lenge forvaltningen utfører den vedtatte politikken på en upartisk måte (Rothstein 2011, Rosanvallon 2011). Denne weberianske forestillingen om et byråkrati som ikke skal ha noen selvstendig innflytelse på politikken har vært hovedskildringen av hvordan den norske politisk-administrative systemet skal være og er.

Samtidig er forskning innen psykologi tydelig på at alle mennesker har systematiske bias som i større eller mindre grad former deres holdninger og atferd. Det er ikke realistisk å anta at saksbehandlere fullt og helt klarer å legge fra seg egne bias i sitt arbeid, selv ikke i profesjoner hvor objektivitet etterstrebes. En måte å utlikne bias er å sørge for at saksbehandlernes bakgrunn reflekterer befolkningen. I det representative byråkratiet skal forvaltningsstaben utgjøre et tverrsnitt av det folket den skal tjene (Lægreid og Olsen 1978; Christensen, Lægreid og Zuna 2001). Utgangspunktet er at den sosiale bakgrunnen til den enkelte tjenestemannen har gjennomslagskraft overfor hens tenkemåte og handlemåte.

Vi har tidligere sett at mange innbyggere tror at saksbehandlere i NAV lar seg påvirke i noen grad av egne holdninger. Med dette som bakteppe er det grunn til å anta at innbyggerne ønsker at saksbehandlerne deler erfaringsbakgrunn med dem selv, slik at de forstår deres situasjon kanskje bedre enn en saksbehandler

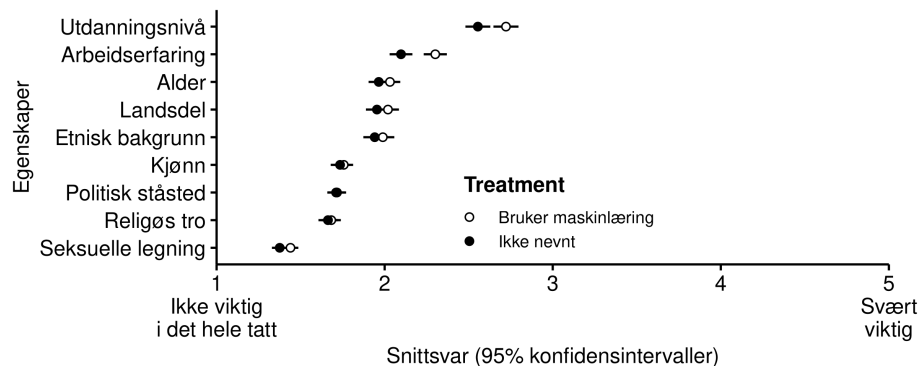
som har en helt annen bakgrunn.

Vi undersøker dette med utgangspunkt i en et design hentet fra en studie om deskriptiv representasjon i politiske beslutningsprosesser. Hvilke egenskaper – om noen – ønsker innbyggerne at saksbehandlerne deler med dem? Figuren under viser at deskriptiv representasjon generelt er lite viktig for innbyggerne. Minst viktig er seksuell legning, mens arbeidserfaring og utdanningsnivå troner øverst.

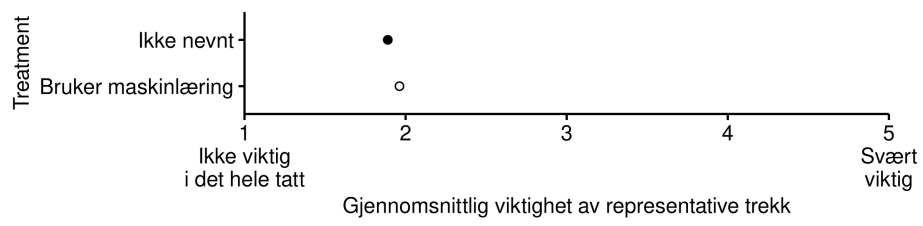
Med vårt fokus på maskinlæring og kunstig intelligens ønsker vi å vite om behovet for deskriptiv representasjon i byråkratiet påvirkes når forvaltningen i tar i bruk dette verktøyet. Det er ikke åpenbart på forhånd hvordan denne utviklingen vil slå ut. På den ene siden kan behovet for at saksbehandlerne deler ens sosiale bakgrunn bli mindre viktig, ettersom alle beslutninger blir mer strømlinjeformede og dermed mindre påvirket av saksbehandlerens bakgrunn.

På den andre siden kan innbyggerne oppleve at man med denne strømlinjeformingen går glipp av viktige nyanser i hver enkelt avgjørelse, og at det er nettopp i slike situasjoner at man er avhengig av saksbehandlere som har forståelse for innbyggernes situasjon og kan gå inn og korrigere i enkelttilfeller.

Resultatene fra eksperimentet viser at folk jevnt over blir mer opptatt av deskriptiv representasjon når forvaltningen benytter seg av maskinlæring og kunstig intelligens som beslutningsstøtte. Vi velger å tolke dette som at kunstig intelligens fører til ytterligere fremmedgjøring av beslutningsprosesser i forvaltningen, som igjen presser fram et følt behov for å ha noen beslutningstakere som kjenner deres situasjon og kan ivarete deres rettigheter og interesser i denne prosessen.



Som det framgår av figuren under er endringene i retning av at deskriptiv representasjon blir viktigere med innføring av maskinlæring likevel små. Deskriptiv representasjon i forvaltning er lite viktig for folk flest, og blir bare litt viktigere ved innføring av maskinlæring og kunstig intelligens.



Bibliography

Barocas, S. and Selbst, A. D. (2016). Big data’s disparate impact.

Chouldechova, A., Benavides-Prado, D., Fialko, O., and Vaithianathan, R. (2018). A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions. In *Conference on Fairness, Accountability and Transparency*, pages 134–148.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Zarsky, T. (2016). The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making. *Science, Technology, & Human Values*, 41(1):118–132.