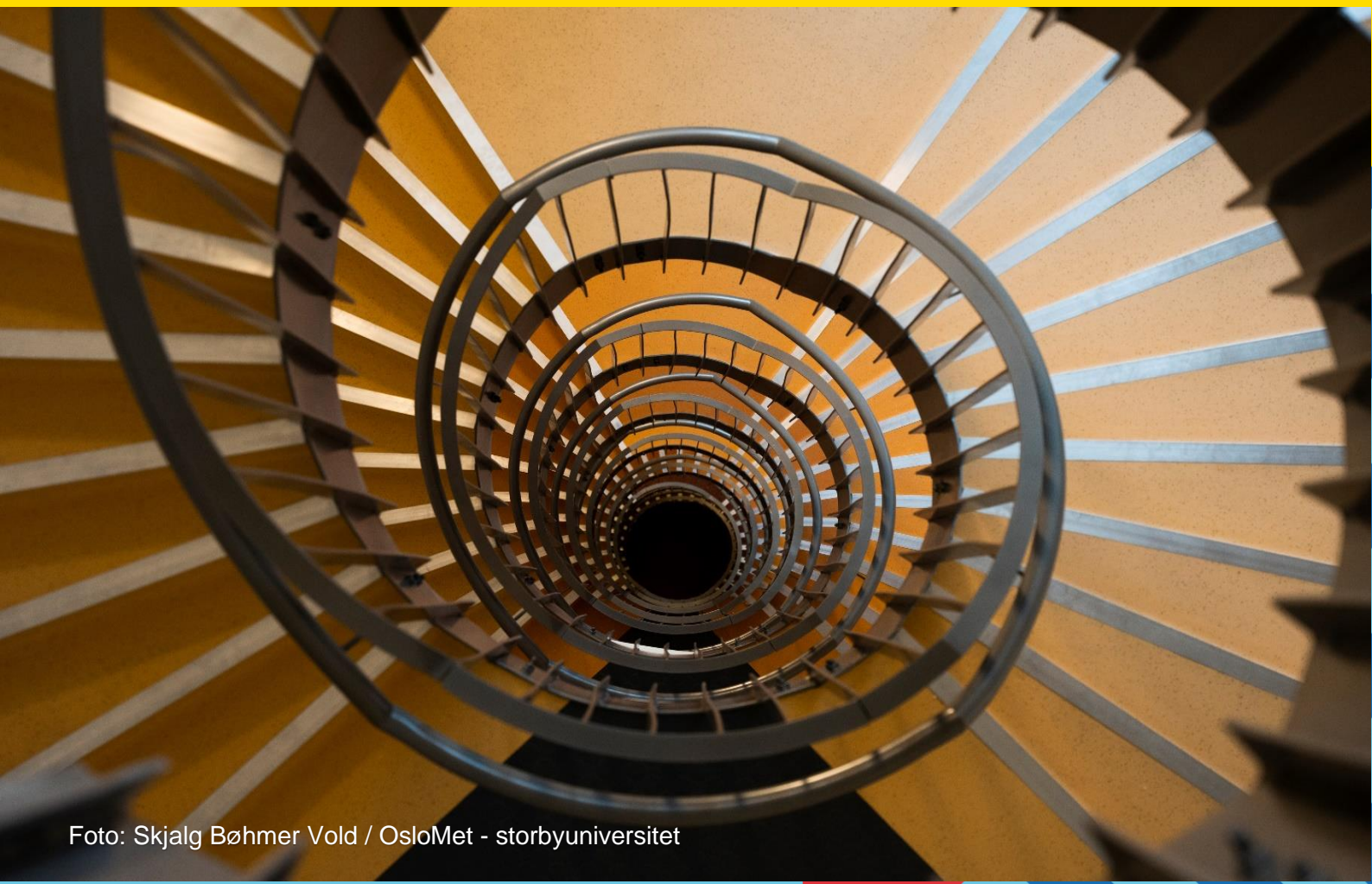


Kartlegging av omfanget av hatefulle og diskriminerende ytringer mot muslimer

av Yuri Kasahara, Tone Liodden, Anis Yazidi, Stian Lid, Ramesh Upreti, Asieh Abolpour Mofrad og Anne Birgitta Nilsen



Kartlegging av omfanget av hatefulle og diskriminerende ytringer mot muslimer

En kartlegging av hatefulle ytringer mot muslimer i digital medier

- Rapporten er utarbeidet på oppdrag av barne-, ungdoms-, og familiedirektoratet (Bufdir)

Følgende enheter ved OsloMet har samarbeidet om kartlegging, analyse og rapport:

- Fakultet for lærerutdanning og internasjonale studier (LUI) ved Institutt for internasjonale studier og tolkeutdanning (IST)
- By- og regionforskningsinstituttet – NIBR
- Fakultet for teknologi, kunst og design (TKD) ved Institutt for informasjonsteknologi (IT)

Prosjektgruppen ønsker å rette en stor takk til; Magnus Klausen, tidligere vit.ass. på Institutt for informasjonsteknologi på OsloMet for arbeidet med å automatisere innhenting av data, Henrik Wiig, professor II ved NIBR for gjennomlesning av tekstutkast og kommentarer, Universitetsbiblioteket på OsloMet for bidrag til kunnskapsoppsummeringen og deltakerne i referansegruppen for gode innspill og kommentarer til rapportutkastet.

ISBN 978-82-8364-397-8 (trykt)

ISBN 978-82-8364-398-5 (online)

Sammendrag

I denne rapporten presenterer vi en studie hvor vi har kartlagt innholdet i og omfanget av ytringer av hatefull og diskriminerende karakter mot muslimer på norske sosiale medier. Studien er gjort på oppdrag fra Barne-, ungdoms- og familiedirektoratet.

I denne studien har vi brukt ulike analysemetoder for å besvare følgende forskningsspørsmål:

1. Hva vet vi om offentlig fremsatte hatefulle og diskriminerende ytringer rettet mot muslimer?
2. Hva inneholder hatefulle og diskriminerende ytringer mot muslimer som fremsettes i norsk offentlighet?
3. Hvor ofte forekommer hatefulle og diskriminerende ytringer mot muslimer?
4. Hvilke muligheter ligger i stordataanalyse og maskinlæring for å kartlegge og kategorisere omfanget av hatytringer nå og fremover?

De to første spørsmålene har vi besvart ved hjelp av en kunnskapsoppsummering av eksisterende kunnskap på feltet, samt noen kvantitative og kvalitative tekstanalyser av hatefulle ytringer. Spørsmål 3–4 har vi svart på gjennom automatisert innhenting av egne stordata og maskinlæring for å klassifisere ytringene.

Kunnskapsoppsummeringen viser at mesteparten av forskningen er fra andre land og hovedsakelig europeiske land; det finnes begrenset med nordisk litteratur om temaet. Vi finner følgende typer studier av hatefulle ytringer mot muslimer:

- stordatastudier, hvor det benyttes automatiserte metoder for å samle inn store mengder data fra sosiale medier
- en god del rent kvalitative studier som tar for seg innholdet i ytringene, gjerne gjennom å skape typologier

Trekk som går igjen på tvers av studiene er fremstillingen av muslimer som iboende voldelige og farlige, en tett kobling mellom muslimer generelt og terrorisme, en forestilling om at muslimer er i ferd med å ta over Europa, kritikk av den politiske venstresidas manglende vilje til å ta inn over seg trusselen fra islam, og fremstilling av kulturelle og verdimeslige forskjeller mellom «islam» og «Vesten» som grunnleggende og uoverstigelige.

Det er færre studier som undersøker

- nettverk og spredning av hatefulle ytringer
- avsendere av hatefulle ytringer
- hvordan hatefulle ytringer på internett påvirker muslimer

Stordataanalysene basert på data fra 2010–2021 viser at

- hatefulle eller diskriminerende ytringer mot muslimer utgjør en liten andel av alle ytringer på norske sosiale medier, men det er en markant økning i tallet på disse
- det er en svært liten gruppe som står bak de fleste av disse ytringene
- I tekstene vi har analysert anses muslimer og Islam som trusler mot Norge, Europa og vestlige verdier.

De kvalitative tekstanalysene viser at de islamkritiske organisasjonene Norge fritt for islam og Stopp islamiseringen av Norge fremmer hatefulle og diskriminerende ytringer mot muslimer.

Summary

In this report, we present a study in which we have mapped both the content and quantity of hate speech towards Muslims on Norwegian social media. The study was commissioned by the Norwegian Directorate for Children, Youth and Families.

In this study, we have used a variety of analysis methods to answer the following research questions:

1. What do we know about publicly made hate speech towards Muslims?
2. What is the content of hate speech made in the Norwegian public sphere towards Muslims?
3. How often does hate speech towards Muslims occur?
4. What opportunities lie in big data analysis and machine learning for the mapping and categorisation of the extent of hate speech now and in the future?

We have answered the first two questions by means of a literature review, as well as some quantitative and qualitative text analyses of hate speech. We have answered the third and fourth questions by way of the automated collection of our own big data and machine learning to classify the statements.

The literature review shows that most research originates from other, mainly European countries, with limited Nordic literature on the subject. We find the following types of studies of hate speech against Muslims:

- Big data studies, where automated methods are used to collect large amounts of data from social media.
- Qualitative studies that address the content of the statements, preferably through the creation of typologies.

Traits that recur across the studies are the portrayal of Muslims as inherently violent and dangerous, the linking of Muslims in general with terrorism, the notion that Muslims are about to take over Europe, the criticism of the political left's unwillingness to address the threat of Islam, and the portrayal of differences in "Islamic" and "Western" cultures and values as fundamental and insurmountable.

There are fewer studies examining:

- networking and the spreading of hate speech,
- the senders of hate speech, and
- how hate speech on the internet affects Muslims.

The big data analyses based on data from 2010 to 2021 show that:

- hate speech towards Muslims makes up a small but sharply increasing proportion of all Norwegian social media,
- there is a very small group behind most of these utterances, and
- in the texts we have analysed, Muslims and Islam are considered threats to Norway, Europe and Western values.

The qualitative text analyses show that the Islam-critical organisations "Norway free of Islam" and "Stop the Islamisation of Norway" promote hate speech towards Muslims.

Innhold

Sammendrag	3
Summary	4
1 Innledning	7
1.1 Bakgrunn.....	7
1.2 Definisjon av hatefulle og diskriminerende ytringer	8
1.3 Data og analyser.....	10
2 Forskningsetiske vurderinger	10
3 Kunnskapsoppsummering	11
3.1 Om litteratursøket	11
3.2 Definisjon av hatefulle ytringer	11
3.3 Problemforståelse.....	13
3.4 Nordisk kontekst	13
3.5 Trekk ved sosiale medier som påvirker ytringene.....	15
3.6 Utløsende faktorer og sammenhegen mellom hendelser på nettet og utenfor nettet	15
3.7 Avsendere av hatefulle ytringer og spredning i nettverk.....	17
3.8 Høyreekstreme partiers rolle i hatefulle ytringer mot muslimer på nettet	19
3.9 Innholdet i hatefulle ytringer	20
3.10 Skille mellom ulike former for islamofobi.....	22
3.11 Ytringer som imøtegår hatefulle ytringer.....	23
3.12 Stordatastudier	25
3.13 Oppsummering.....	27
4 Analyser av omfang – stordataanalyser	28
4.1 Metode	28
4.2 Analyser – Omfang av omtale av muslimer	35
4.3 Oppsummering	41
5 Analyser av innholdet i hatefulle og diskriminerende ytringer	42
5.1 Kvantitative analyser.....	42
5.2 Kvalitative analyser.....	45

6 Konklusjon	50
Referanser	51
Vedlegg - Søkehistorikk.....	57
Problemstilling, avgrensninger	57
Valg av databaser.....	57
Skandinaviske databaser:	58

Figurer

Figur 1 – Totalt antall nyhetssaker på nyhetsmediers Facebook-sider per måned – alle medier	36
Figur 2 – Antall nyhetssaker på Twitter per måned – alle medier	37
Figur 3 – Totalt antall brukeraktiviteter på Facebook per måned – alle medier	38
Figur 4 – Antall brukeraktivitet på Twitter per måned – alle medier	38
Figur 5 – Andel i prosent av hatefulle kommentarer på Twitter	40
Figur 6 – Andel i prosent av hatefulle kommentarer på utvalg av Facebook-sider	41
Figur 7 – Ordsky fra Twitter 30.11.21 (Datasett 3).....	43
Figur 8 – Ordsky fra Facebook 30.11.21 (Datasett 4)	43
Figur 9 – Plakater fra Norge fritt for islam sin nettside.....	46
Figur 10 – Plakat fra SIAN sin nettside	47
Figur 11 – Plakat fra SIAN sin Facebook-side	47
Figur 12 – Plakat fra SIAN sin Facebook-side	48
Figur 13 – Plakat fra SIAN sin Facebook-side	49
Figur 14 – Tekstfil fra SIAN sin Facebook-side.....	49

Tabeller

Tabell 1 – Utfallsrom – relasjon mellom predikerte og faktiske klassifiseringer	31
Tabell 2 – Modellens ytelse på testsett.....	32
Tabell 3 – Modellens ytelse på uklassifiserte data.....	33
Tabell 4 – Tema-modellering for Twitter.....	44
Tabell 5 – Tema-modellering for Facebook.....	44

1 Innledning

Utbredelsen av internett og bruken av sosiale medier har gjort det vesentlig lettere for folk flest å ytre seg offentlig. Digitaliseringen har slik lagt til rette for økt deltagelse i samfunnsdebatten, og digitale plattformer har blitt en stadig viktigere arena for det offentlige ordskiftet. Terskelen for å fremsette hatefulle og diskriminerende ytringer ser samtidig ut til å ha blitt lavere ved bruken av internett og sosiale medier, og de fremsatte ytringene kan spres raskere og nå videre enn noensinne (Evolvi, 2018). Dette kan være på plattformer som Facebook, Twitter, YouTube, kommentarfelter på nettaviser og andre nettsteder. Digitale medier er dermed antagelig den største arenaen for hatefulle og diskriminerende ytringer. En viktig årsak er muligheten for å opptre anonymt, fordi man på den måten unngår å måtte stå til ansvar for sine ytringer (Dechesne et al., 2020). Noen hatefulle ytringer faller inn under straffelovens bestemmelser, og ved å opptre anonymt kan man lettere unngå straffeforfølgelse fordi det kan være vanskelig for politiet å finne ut av hvem som står bak ytringene. Hatefulle ytringer er dessuten en sentral faktor i radikaliserings og ekstremisme (Hassan et al., 2018; Baldauf et al., 2019). Det er velkjent at både Filip Manshaus¹ og Anders Behring Breivik (Ravndal, 2013) deltok i samtaler preget av hatefulle og diskriminerende ytringer på ulike sosiale plattformer.

1.1 Bakgrunn

Bakgrunnen for denne rapporten er Barne-, ungdoms- og familiedirektoratet (Bufdir) sin utlysning av et FoU-oppdrag om å kartlegge innholdet i og omfanget av ytringer av hatefull eller diskriminerende karakter mot muslimer. Det utlyste FoU-oppdraget er en oppfølging av tiltak 17 i Regjeringens handlingsplan mot rasisme og diskriminering på grunn av etnisitet og religion (2020–2023). Oppdraget skal også ses i sammenheng med Regjeringens handlingsplan mot diskriminering av og hat mot muslimer (2020–2023) og Regjeringens strategi mot hatefulle ytringer (2016–2020). Formålet med FoU-oppdraget er å få oversikt over innholdet i og omfanget av hatefulle eller diskriminerende ytringer mot muslimer som forekommer på norskspråklige sosiale medier, som foreslått i en tidligere studie utført av Marjan Nadim og Audun Fladmoe (2016). Oppdraget er løst av OsloMet – storbyuniversitetet ved Fakultet for lærerutdanning og internasjonale studier (LUI), Fakultet for teknologi, kunst og design (TKD) og By- og regionforskningsinstituttet NIBR. Prosjektgruppen er satt sammen av forskere ved OsloMet med særlig kompetanse på hatytringer og bruk av stordata. Anne Birgitta Nilsen, professor i internasjonale studier ved Fakultet for lærerutdanning og internasjonale studier, har ledet prosjektet. Professor Anis Yazidi; sammen med førstelektor Asieh Abolpour Mofrad og masterstudent i data science Ramesh Upreti; ved Institutt for informasjonsteknologi og kunstig intelligens bidro i forhold til utviklingen av dyplæringsmodellen «Natural Language Processing». Yuri Kasahara, Tone Maia Liodden og Stian Lid er forskere ved By- og regionforskningsinstituttet NIBR, OsloMet og bidro bredt inn i prosjektet med hensyn på kunnskapsgrunnlag, ordlister, klassifisering av data, bruk av stordata (utvikling og revidering av algoritmer), analyse og rapportskrivning.

I forbindelse med prosjektet ble det opprettet en referansegruppe bestående av representanter med særlig kompetanse på hatytringer mot muslimer, der følgende organisasjoner har bidratt med innspill, forslag og kommentarer i forbindelse med rapporten: Islamsk Råd Norge, Antirasistisk senter, Sekretariatet til Ytringsfrihetskommisjonen, Kripos, LDO, Minotenk, Agenda X, Stopp

¹ <https://www.dagbladet.no/kultur/en-farlig-forenkling/72548421>

Hatprat, Islamic Cultural Center, Salam Norge, Norsk folkehjelp og Stiftelsen 10. august. Referansegruppen har også bidratt i våre diskusjoner av fenomenet hatefulle og diskriminerende ytringer.

1.2 Definisjon av hatefulle og diskriminerende ytringer

Hatefulle og diskriminerende ytringer er omstridte begreper, og det finnes ingen enighet om hvordan disse skal defineres eller operasjonaliseres. Hatefulle og diskriminerende ytringer er knyttet til straffeloven §185. Det er straffbart å fremme grove hatefulle eller diskriminerende ytringer mot enkeltpersoner. I straffeloven er en hateful ytring definert som det å true eller forhåne noen, eller fremme hat, forfølgelse eller ringeakt overfor noen, på bakgrunn av bestemte diskrimineringsgrunnlag. Straffeloven verner mot diskriminerende og hatefulle ytringer som fremsettes offentlig eller i andres nærvær på grunn av noens: a) hudfarge eller nasjonale eller etniske opprinnelse, b) religion eller livssyn, c) seksuelle orientering eller d) nedsatte funksjonsevne. Det er også vedtatt at transpersoner og andre som har en kjønnsidentitet eller et kjønnsuttrykk som bryter med omgivelsenes forventninger, skal gis et vern mot hatefulle ytringer.² Ytringsfriheten i Grunnloven §100 er altså ikke absolutt, men om en ytring er straffbar eller ikke, kommer an på sammenhengen ytringen er fremsatt i. Det er mange ytringer som kan oppleves som både ubehagelige og sårende, men som ikke er straffbare. For å vurdere om en ytring er straffbar, må ytringen tolkes i den sammenheng den ble ytret, slik det for eksempel beskrives i en dom fra høyesterett i 2018.³ Vilde Jalleni Tennfjord og Erlend Andreas Methi fremhever at:

Hvor grensen skal gå for hvilken menneskerettighet som veier tyngst – ytringsfriheten eller diskrimineringsvernet – og følgelig hvor grensen går for når en ytring er straffbar, er et tilbakevendende spørsmål både i samfunnsdebatten og i rettsapparatet.⁴

Dette prosjektet er imidlertid ikke begrenset til straffelovens definisjon og operasjonalisering av hatefulle og diskriminerende ytringer. Vi har anlagt et videre perspektiv på dette språklige fenomenet – et fenomen som gjerne omtales som hatprat og som omfatter verbale, ikke-verbale, skriftlige og visuelle uttrykk som beskrevet av Nilsen (2014b). Slike bredere definisjoner er også anlagt av den europeiske kommisjon mot rasisme og intoleranse (ECRI) og Likestillings- og diskrimineringsombudet (LDO). ECRI knytter hatprat til rasisme og diskriminering.⁵ Likestillings- og diskrimineringsombudet (LDO) bygger på ECRI (2015) og Nilsen (2014b) i sin definisjon av hatefulle ytringer (LDO, 2015, s. 12). Samtidig erkjenner vi at betegnelsen hatprat kan oppfattes som misvisende, for i denne betegnelsen ligger det en antagelse om dem som ytrer seg på en slik måte – at de hater og/eller at de har til hensikt å spre hat. Slike antagelser bør vi være varsomme med. Det vi gjerne omtaler som hatprat kan like gjerne være styrt av frykt, misunnelse eller sinne. Bak de hatefulle ytringene kan det også stå såkalte nettroll som morer seg med å provosere. Folk har ulik motivasjon for å delta i opphetede og aggressive nettdebatter (Nadim, Thorbjørnsrud og Fladmoe, 2021).

Et viktig kjennetegn ved hatprat er at det er godt egnet til å svekke et individ eller en gruppe menneskers status og anseelse i samfunnet (Waldron, 2012). Dette språklige fenomenet kunne vi

² Prop. 66 L (2019–2020) kap. 21.4

³ <https://www.domstol.no/globalassets/upload/hret/avgjorelser/2018/avgjorelser-april-2018/saknr-2018-34-anonymisert.pdf>

⁴ <https://juridika.no/innsikt/hatefulle-ytringer-grensen-mellom-ytringsfrihet-og-diskrimineringsvern-kommentar-til-hr-2020-2133-a>

⁵ <https://www.coe.int/en/web/european-commission-against-racism-and-intolerance/recommendation-no.15>

derfor også kalt for nedsettende ytringer. En fordel ved en slik betegnelse er at nedsettende ytringer er noe vi alle kan uttale oss om. Vi vet alle hvilke ytringer som kan være egnet til å svekke et individ eller en gruppe menneskers status og anseelse i samfunnet. Et annet typisk trekk ved dette språklige fenomenet er at det sikter mot sine lesere og tilhøreres negative følelser, og ikke mot fornuften. Disse ytringene har derfor et potensial for å vekke eller forsterke negative følelser, holdninger og oppfatninger hos tilhørerne. I dette tilfellet negative følelser, holdninger og oppfatninger overfor muslimer. Men saken er mer komplisert enn som så, for muslimer kan oppleve seg hatet når de blir møtt med dette språklige fenomenet. Denne opplevelsen tilsier at hatefulle ytringer eller hatprat er treffende betegnelser. Men så er det også slik at muslimer kan føle seg mislikt, mistenkeliggjort eller uønsket og ikke nødvendigvis hatet.

Hatefulle ytringer kan påvirke både dem som er vitner, lesere og tilhørere og dem som rammes av hatefulle ytringer. Det er dessuten slik at hatefulle ytringer ofte forekommer og spres i forum og på steder der de som omtales i ytringene ikke er til stede. Det kan være blant venner på nett eller andre steder. Hatefulle ytringer har også konsekvenser for samfunnet, fordi de kan skape splid. Det finnes imidlertid lite kunnskap om hvordan hatefulle ytringer henger sammen med negative holdninger overfor muslimer i befolkningen generelt.

Denne rapporten omhandler sosiale medier og er, som nevnt, gjort på oppdrag fra Barne-, ungdoms- og familiedirektoratet, og vi vil derfor bruke betegnelsen som lå i oppdraget: hatefulle og diskriminerende ytringer, men presisere at dette prosjektet handler om et større språklig fenomen som også drives av andre følelser enn hat og av andre intensjoner enn å spre hat. Fenomenet har også en rekke andre kjennetegn, som beskrevet av Nilsen (2014b og 2022). Det er dessuten en nær kobling mellom holdninger, stereotyper og fordommer og hatefulle og diskriminerende ytringer, da disse spres ved hjelp av hatprat. En undersøkelse utført av HL-senteret viser at negative stereotyper om muslimer er utbredt i det norske samfunnet.

Samlet sett viser 34,1 prosent av befolkningen utpregede fordommer mot muslimer. Hele 47 prosent av respondentene støtter påstanden «Muslimer har selv mye av skylden for økende muslimhets»; påstanden «Muslimer vil ikke integreres i det norske samfunnet» støttes av 42 prosent; 39 prosent støtter at «Muslimer utgjør en trussel mot norsk kultur», og 30 prosent mener at «Muslimer ønsker å ta over Europa». Forholdsvis mange uttrykker også negative følelser og sosial avstand overfor muslimer. Det er 27,8 prosent som føler motvilje mot muslimer, og samlet sett 19,6 prosent som vil mislike å ha muslimer som naboer eller i vennekretsen. (Hoffmann & Moe, 2017, s. 7–8)

I hatefulle og diskriminerende ytringer mot muslimer gjøres muslimene kollektivt ansvarlige for terror fra ekstreme islamister, for politikk i andre land der det bor muslimer og for et påstått forfall i Norge og Europa (Evolvi, 2019; Cleland et al., 2018; Brindle, 2016; Oboler, 2016; Ekman, 2015). Her er det ingen henvisninger til konkrete grupper, organisasjoner eller lignende. Det fremmes også påstander om at muslimer har en særegen mentalitet som ikke lar seg forene med såkalte vestlige verdier eller demokratiet. I den høyreekstreme konspirasjonsteorien om Eurabia ligger det forestillinger om at muslimer i skjul arbeider for en islamisering eller kulturell dominans av Norge og Europa i samarbeid med enkelte politiske partier og myndigheter (Bangstad, 2014; Oboler, 2016; Ekman, 2015). Islams fremtidige dominans i Vesten blir muliggjort av den politiske venstresida og liberale politikere, som i praksis understøtter muslimenes plan om å få herredømme over Vesten (Gaudette et al., 2020; Oboler, 2019; Cleland et al., 2018; Brindle, 2016; Ekman, 2015).

I kunnskapsoppsummeringen under har vi en nærmere beskrivelse av både hatefulle ytringer generelt og innholdet i hatefulle ytringer mot muslimer spesielt. I metodedelens beskriver vi hvordan vi har operasjonalisert begrepet hatefulle ytringer mot muslimer

1.3 Data og analyser

I denne studien har vi brukt ulike analysemetoder for å besvare følgende forskningsspørsmål:

- Hva vet vi om offentlig fremsatte hatytringer rettet mot muslimer?
- Hva inneholder hatytringer mot muslimer som fremsettes i norsk offentlighet?
- Hvor ofte forekommer hatytringene?
- Hvem fremsetter hatytringene?
- Hvilke muligheter ligger i stordataanalyse og maskinlæring for å kartlegge og kategorisere omfanget av hatytringer nå og fremover?

Det første spørsmålet har vi besvart ved hjelp av en kunnskapsoppsummering av eksisterende kunnskap på feltet. Spørsmål 2–5 har vi løst gjennom innhenting av egne stordata. Hatefulle ytringer på nettet inkluderer imidlertid ikke bare skriftlige tekster, men også bilder, symboler og andre ikke-tekstlige ressurser, såkalte sammensatte eller multimodale ytringer (jfr. Nilsen, 2012; Kress, 2010). Vi har derfor gjort kvalitative analyser for å gi noen inntrykk av hvordan disse multimodale (kombinasjoner av skrift, bilder og symboler) hatefulle ytringene mot muslimer manifesterer seg på nettet.

Stordata er en analysemetode som innebærer innhenting og analyse av data som er så store at de krever mer enn vanlig datakraft for å samles inn, lagres og analyseres. Våre datasett inneholder 800 000 kommentarer om muslimer fra Facebook og Twitter for perioden januar 2010 til september 2021. Vi har utviklet en statistisk modell og trent denne til å gjenkjenne og finne hatefulle ytringer, altså å skille hatefulle ytringer fra ikke-hatefulle ytringer. Den modellen vi har laget for å gjenkjenne hatefulle og ikke-hatefulle ytringer anser vi for å være en god modell, fordi den i svært liten grad overestimerer antallet hatefulle ytringer, og fordi den er i stand til å fange opp forskjeller mellom hatefulle og ikke-hatefulle ytringer over tid. Den er også den første norskspråklige modellen laget for dette formålet, og sammenlignet med gode internasjonale modeller er vår modell minst like god.

Gjennom prosessen har vi systematisk dokumentert våre erfaringer med å bruke stordata på hatefulle og diskriminerende ytringer. Med denne rapporten håper vi derfor ikke å bare bidra med ny og nyttig kunnskap om diskriminerende og hatefulle ytringer mot muslimer, men også med kunnskap om hvordan man kan bruke stordata og maskinlæringsmetoder for å videreutvikle forskningen på feltet.

Metode blir beskrevet mer detaljert underveis i rapporten under de ulike metodene.

2 Forskningsetiske vurderinger

En stordataanalyse av enkeltpersoners ytringer fører med seg en rekke etiske spørsmål. Prosjektet er meldt til NSD, i henhold til lov om personvern, og er gjennomført i tråd med de forskningsetiske retningslinjene fra Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora (NESH). Særlig viktig for denne studien er å beskytte identiteten til enkeltindivider som har fremsatt ytringer på sosiale medier, selv om vi kun har samlet inn data publisert på offentlige medier. Alle navn er anonymisert ved hjelp av en indeksering. Unntak er organisasjoner eller enkeltindivider som regnes som en offentlig person (for eksempel politikere eller ledere av organisasjoner).

Vi har også valgt å gjøre noen kvalitative analyser av hatefulle og diskriminerende ytringer i noen sammensatte tekster. Disse tekstene har vi valgt å gjengi i kapittelet om analyser av innholdet i hatefulle og diskriminerende ytringer, for å vise hvordan disse kan se ut og hvilken virkning de potensielt kan ha. En innvending kan være at vi på den måten bidrar til å spre hatefulle og

diskriminerende ytringer. Vi tror ikke det, ettersom denne rapporten trolig ikke vil leses av folk som har interesse av å spre disse tekstene.

3 Kunnskapsoppsummering

3.1 Om litteratursøket

Ved hjelp av bibliotekar ved OsloMet har det blitt gjennomført et strukturert søk for å identifisere forskningslitteratur og rapporter produsert av frivillige organisasjoner og andre aktører om hatefulle ytringer mot muslimer på internett.⁶ Søket er avgrenset til tidsperioden 2010–2020. Det har blitt søkt på engelsk, norsk, svensk og dansk i internasjonale og skandinaviske databaser. Hovedtemaer i søket har vært ytringer – muslimer – nettbasert (se vedlegg for mer oversikt over søkeord og mer informasjon om søket). Etter at potensiell relevant litteratur ble identifisert gjennom søkemotorer, gikk vi gjennom alle bidragene, blant annet ved å lese sammendrag, for å skille ut bidrag som handlet om vårt tema. I tillegg til det strukturerte søket har vi supplert noe med manuelle søk i Google Scholar (hate speech AND muslims AND online) og ved å ta med relevante artikler og rapporter som har dukket opp i litteraturlisten til artikler som ble fanget opp i søket. Til sammen baserer denne kunnskapsoppsummeringen seg på 68 forskningsbidrag. Vi har gruppert litteraturen inn i ulike temaer som vi beskriver under.

I litteraturgjennomgangen har vi begrenset oss til artikler hvor søkeord knyttet til islam eller muslimer opptrer i sammendrag eller i tittel, eller hvor det tydelig fremkommer at disse temaene er viktige i artikkelens hovedtekst. Det er likevel verdt å merke seg at temaer knyttet til islam og muslimer også dukker opp i flere studier som undersøker hatefulle ytringer generelt, og som derfor ikke er inkludert her. Når temaene «islam» og «muslimer» dukker opp i såpass mye litteratur som handler om hatefulle ytringer mer overordnet, kan det anses som en indikasjon på at muslimer er en gruppe som er spesielt utsatt for slike ytringer.

3.2 Definisjon av hatefulle ytringer

Litteraturgjennomgangen synliggjør at det ikke finnes en tydelig, omforent definisjon av hatefulle ytringer (se også Nadim et al., 2017, s. 18). En årsak er at hatefulle ytringer ikke er et avgrenset fenomen, men kan forstås som et kontinuum, fra ytringer som er subtile og klassifiseres som mindre alvorlige, til ytringer som er straffbare. Oppfatninger av hva som utgjør hatefulle ytringer blir formet av det enkelte lands historiske, kulturelle og sosiale kontekst – noe som også former lovverkene som regulerer grensene mellom lovlige og ulovlige ytringer (Baider et al., 2020). Hatefulle ytringer har også en rekke grenseflater opp mot andre fenomener, som mobbing, diskriminering og vold (Eggebo & Stubberud, 2016; Nilsen, 2014b). I mange studier benyttes begrepet «hatefulle ytringer» og «islamofobi» eller «islamofobe ytringer» om hverandre. Bruk av begrepet «islamofobi» er i seg

⁶ Merk at et «strukturert» søk ikke er det samme som et «systematisk» litteratursøk. Et systematisk søk ville gitt en mer fullstendig oversikt over litteraturen på feltet. Systematiske litteratursøk innebærer en egen metodikk med blant annet strenge kriterier for hvilke artikler som faller innenfor og utenfor. Et slikt søk ville kreve mer ressurser enn vi har hatt til rådighet i dette prosjektet.

selv kontroversielt, og på samme måte som med hatefulle ytringer, finnes det ingen tydelig enighet om hvordan begrepet bør defineres (Vidgen et al., 2021).

De fleste empiriske undersøkelsene legger til grunn en bred forståelse av hatefulle ytringer som inkluderer både lovlige og ulovlige ytringer. I norsk kontekst innebærer en hatefull ytring i lovmessig forstand å «true eller forhåne noen, eller fremme hat, forfølgelse eller ringeakt overfor noen på grunn av deres hudfarge eller nasjonale eller etniske opprinnelse, religion eller livssyn, seksuelle orientering, kjønnsidentitet eller kjønnsuttrykk eller nedsatte funksjonsevne» (Straffeloven, 2005, § 185). En viktig del av den strafferettslige definisjonen på hatefulle ytringer er at slike ytringer må være fremsatt offentlig eller «i andres nærvær».

I mange av studiene, og særlig dem som benytter stordata for å klassifisere ytringer, ville det være en umulig oppgave å foreta juridiske vurderinger av hvorvidt ytringene faller inn under den lovmessige definisjonen av hatefulle ytringer. I Norge er det dessuten begrenset med rettspraksis på feltet, og det er krevende å spesifisere når en ytring bør anses for å være ulovlig på et generelt grunnlag. Rettsavgjørelser beror ofte på en «konkret helhetsvurdering», hvor også forsett og uaktsomhet skal vurderes som en del av skyldkravet (Veledar & Burkal, 2018, s. 9). Det ville heller ikke være produktivt å fokusere utelukkende på de ulovlige ytringene, fordi lovlige og ulovlige hatefulle ytringer er en del av det samme fenomenet, og overgangene er glidende. Selv om ikke ytringer er ulovlige, kan de ha store konsekvenser for enkeltindivider og for samfunnet (Nadim et al., 2017).

Et fellestrekk i forståelsen av hatefulle ytringer er likevel at slike ytringer knyttes til en persons tilskrevne eller faktiske gruppetilhørighet. Ytringer som er rettet mot rent individuelle egenskaper faller dermed utenfor definisjonen av hatefulle ytringer (Nadim et al., 2017, s. 15) og beskrives gjerne som mobbing (Nilsen, 2014b). I forskningslitteraturen blir hatefulle ytringer ofte beskrevet som ytringer rettet mot karakteristikk som etnisitet, hudfarge, religion, kjønn, funksjonsnedsetting eller seksuell orientering (se f.eks. Baider et al., 2020, s. 172; Oboler, 2016, s. 42). Dette er grunnlag som de fleste europeiske land har forpliktet seg til å beskytte gjennom lovverk mot diskriminering. Noen studier inkluderer også kategorier som politisk oppfatning og utseende (se f.eks. Veledar & Burkal, 2018; Zuleta et al., 2017). En definisjon av hatefulle ytringer som ofte blir sitert i forskningslitteraturen er fra European Commission against Racism and Intolerance (ECRI):

[T]he advocacy, promotion or incitement, in any form, of the denigration, hatred or vilification of a person or group of persons, as well as any harassment, insult, negative stereotyping, stigmatization or threat in respect of such a person or group of persons and the justification of 'race', colour, descent, national or ethnic origin, age, disability, language, religion or belief, sex, gender, gender identity, sexual orientation and other personal characteristics or status. (ECRI, 2016, s. 2)

Fordi ulike studier benytter ulike definisjoner av hatefulle ytringer, vil resultater knyttet til omfang av hatefulle ytringer også variere (Nadim et al., 2017). Mens det er relativt stor enighet om at oppfordring til vold mot en gruppe er en hatefull ytring, vil det ofte være uenighet om andre typer ytringer. I en studie inkluderer forfatterne for eksempel ytringer som tar til orde for restriksjoner i innvandreres rettigheter i sin definisjon av hatefulle ytringer (Calderón et al., 2020). En slik bred definisjon vil det sannsynligvis være knyttet mer uenighet til. Selv innenfor en og samme studie, hvor samme definisjon og kriterier benyttes, vil ulike personer til en viss grad klassifisere samme ytring på ulikt vis, fordi det alltid vil være et subjektivt element i slike vurderinger (se f.eks. Poletto et al., 2017, s. 5; Waseem & Hovy, 2016, s. 90).

3.3 Problemforståelse

Det er flere årsaker til at forskere mener at hatefulle ytringer mot muslimer er et viktig fenomen å studere og forebygge. Flere forskere oppfatter aggresjon mot muslimer på nettet og utenfor nettet som samme fenomen (Laurent, 2020, s. 2147; Awan & Zempi, 2017, s. 106). Hatefulle ytringer kan potensielt være forløperen til annen hatkriminalitet (Awan, 2014, s. 134; Nilsen, 2014b). Forskere peker på at hatefulle ytringer mot muslimer på nettet kan bidra til å normalisere hat mot muslimer (Oboler, 2016), og at bruk av sosiale medier kan forsterke fordommer. I en undersøkelse fra Polen (Soral et al., 2020) fant forskere at personer som primært brukte tradisjonelle medier som sin nyhetskilde, skåret lavere på fordommer mot muslimer enn personer som i stor grad brukte sosiale medier. Selv om forskerne ikke kan etablere en kausalsammenheng – altså at de sosiale mediene i seg selv forårsaker høyere fordommer – kan studien være en indikasjon på at digitale medier bidrar til å forsterke og normalisere negative holdninger mot muslimer og til å endre holdningene til hva slags ytringer som er akseptable. Hatefulle ytringer blir ansett for å være skadelige, ikke bare for de som mottar slike ytringer direkte, men også for andre som observerer ytringene (Nadim et al., 2017, s. 63). En studie av hvordan muslimer opplever islamofobi, tyder på at opplevelsen av å tilhøre en gruppe som er fryktet og utsatt i samfunnet, i seg selv gir negative psykologiske konsekvenser, uavhengig av om man som individ har blitt direkte utsatt for islamofobi eller diskriminering (Kunst et al., 2013). Hatefulle ytringer anses også for å ha en negativ effekt på samfunnet ellers, fordi de skaper sosiale skillelinjer (Nilsen, 2014b). Slike ytringer kan være en viktig ingrediens for å bygge gruppeidentitet og nettverk på ytre høyre-sida, på tvers av landegrensler (se f.eks. Nilsen et al., 2020).

I den norske debatten blir hatefulle ytringer ofte diskutert i sammenheng med ytringsfrihet (Veledar & Burkal, 2018, s. 5). Mens noen mener at tiltak mot hatefulle ytringer kan undergrave ytringsfriheten, mener andre at det er helt nødvendig å bekjempe hatefulle ytringer for å sikre at alle grupper kan delta i den offentlige debatten – og dermed benytte sin ytringsfrihet – uten å bli utsatt for hets og trusler (Lid et al., 2020, s. 56). I den internasjonale litteraturen vi har tatt for oss her, blir imidlertid diskusjoner knyttet til ytringsfrihet og hatefulle ytringer viet overraskende liten plass, utover at noen forskere nevner at det kan være problematisk å bruke lovverk for å bekjempe hatefulle ytringer fordi det kan føre til en innsnevring av ytringsfriheten (se f.eks. Aguilera-Carnerero & Azeez, 2016, s. 23).

3.4 Nordisk kontekst

Det finnes, så vidt vi vet, ingen undersøkelser i de nordiske landene som direkte undersøker hatefulle ytringer mot muslimer på internett, men det har blitt gjennomført noen undersøkelser i Danmark og Norge som tar for seg hatefulle ytringer i sosiale medier mer generelt (Veledar & Burkal, 2018; Zuleta et al., 2017). Forskere ved ISF har laget en kunnskapsoppsummering om hatefulle ytringer på internett, hvor de blant annet tar for seg erfaringer med hatefulle ytringer, konsekvenser og avsenderes karakteristikk og motivasjon (Nadim et al., 2017). I rapporten undersøkes også omfang av hatefulle ytringer i Norge gjennom en spørreundersøkelse. Det fremkommer at sju prosent av respondentene rapporterer at de har mottatt det de selv oppfatter som hatefulle ytringer via sosiale medier. Her er ytringer rettet mot alle typer grunnlag med, inkludert, for eksempel, ytringer rettet mot politiske meninger og utseende. Om man ser på grunnlagene som dekkes i straffeloven (som nasjonalitet og seksuell legning), er det bare to prosent som svarer at de har vært utsatt for hatefulle ytringer (Nadim et al., 2017, s. 103).

Blant grunnlagene som dekkes av straffeloven, blir nasjonalitet og religion oftest oppgitt blant dem som har mottatt hatefulle ytringer (Nadim et al., 2017, s. 97). Personer med innvandrerbakgrunn

opplever ikke å motta hatefulle ytringer oftere enn majoritetsbefolkningen, men slike ytringer rettes oftere mot egenskaper som dekkes av straffeloven, som nasjonalitet og religion, mens majoritetsbefolkningen oftere får kommentarer rettet mot andre grunnlag, slik som politisk ståsted eller utseende (Nadim et al., 2017, s. 116)⁷. Sosiale medier og kommentarfelter er de vanligste arenaene for hatefulle ytringer, etterfulgt av arbeidsplassen og offentlig transport (Fladmoe et al., 2019, s. 35).

I en annen undersøkelse utforsker Likestillings- og diskrimineringsombudet (LDO) (Veledar & Burkal, 2018) hatefulle ytringer på NRK og TV2 sine Facebook-sider, i tillegg til at det ble gjennomført en brukerundersøkelse om erfaringer med hatefulle ytringer på sosiale medier. En tilsvarende studie har blitt gjennomført i Danmark, på Facebook-sidene til Danmarks Radio og TV Nyhederne (Zuleta et al., 2017). Undersøkelsene ser på hatefulle ytringer generelt, ikke ytringer spesielt rettet mot muslimer. I sin undersøkelse klassifiserer LDO ca. sju prosent av de innhentede kommentarene som hatefulle (Veledar & Burkal, 2018, s. 38). I Danmark var tallet mer enn dobbelt så høyt, på rundt 15 prosent (Zuleta et al., 2017, s. 59).

LDO finner at kommentarer med hatefulle ytringer skaper mer aktivitet enn ikke-hatefulle ytringer (Veledar & Burkal, 2018, s. 55). I begge studiene finner man at det er menn som skriver flest hatefulle ytringer (Veledar & Burkal, 2018, s. 47; Zuleta et al., 2017, s. 8), og at hovedvekten av hatefulle ytringer retter seg mot grupper, snarere enn enkeltpersoner (Veledar & Burkal, 2018, s. 50; Zuleta et al., 2017, s. 71). Flest hatefulle ytringer rettes mot personers eller gruppers politiske syn – dette utgjør 53 prosent av de hatefulle ytringene i Norge og 31 prosent i Danmark (Veledar & Burkal, 2018, s. 51; Zuleta et al., 2017, s. 72). Deretter er det etnisitet som er det viktigste grunnlaget ytringene rettes mot, i 19 prosent av tilfellene i Norge og 22 prosent av tilfellene i Danmark (Veledar & Burkal, 2018, s. 51; Zuleta et al., 2017, s. 72). Nesten like viktig er religion: 18 prosent av de hatefulle ytringene i LDOs undersøkelse var rettet mot religion eller livssyn (Veledar & Burkal, 2018, s. 51); i Danmark var tilsvarende tall 21 prosent (Zuleta et al., 2017, s. 72). Av disse var hele 88 prosent rettet mot islam i de norske mediene (Veledar & Burkal, 2018, s. 54), og 86 prosent var rettet mot islam i de danske mediene (Zuleta et al., 2017, s. 74). Fire temaer utløser flest hatefulle ytringer på Facebook-sidene til TV2 og NRK: flyktninger og integrering, krig og katastrofer, likestilling og diskriminering og utenriksnyheter (Veledar & Burkal, 2018, s. 44). I Danmark er det lignende temaer som utløser hatefulle ytringer, nemlig religion og tro, flyktninger, migranter og asyl og likestilling (Zuleta et al., 2017, s. 59).

LDO fulgte opp undersøkelsen fra 2018 med en studie hvor de også tok spesifikt for seg hatefulle ytringer på Facebook (Øistad & Veledar, 2021). I undersøkelsen finner man at mange synes tonen i debattene på Facebook er såpass hard at de lar være å delta i debatter: 69 prosent av kvinnene i undersøkelsen vegrer seg for å delta i debatter på Facebook på grunn av tonen, mens 55 prosent av de mannlige respondentene svarer det samme (Øistad & Veledar, 2021, s. 42). I tråd med funnene i forrige rapport lar flest være å debattere temaer som utlendinger, flyktninger, innvandrere, religion og tro (Øistad & Veledar, 2021, s. 42). I denne undersøkelsen var det 14 prosent som oppga at de mente den hatefulle ytringen som hadde blitt rettet mot dem, handlet om deres religiøse bakgrunn (Øistad & Veledar, 2021, s. 51)

Politiet rapporterer årlig om hatkriminalitet i Norge, herunder hatefulle ytringer. I 2020 hadde elleve prosent av hatkriminaliteter som ble anmeldt til politiet, sosiale medier som «åsted». Politiet presiserer imidlertid at det antagelig er store mørketall her, både fordi publikum er mindre tilbøyelige til å anmelde hatkriminalitet på sosiale medier og fordi Oslo politidistrikt ikke selv har opprettet så mange slike saker (Oslo politidistrikt, 2020, s. 21).

⁷ Se også Staksrud et al. (2014) som delvis benytter samme datagrunnlag.

3.5 Trekk ved sosiale medier som påvirker ytringene

Det som gjør sosiale medier til en unik arena for spredning av hatefulle ytringer, er hvor raskt det går å spre slike ytringer (Evolvi, 2018). Kjernen i islamofobe narrativer har ikke nødvendigvis endret seg så mye de siste årene, men teknologien og logikken i sosiale medier gir dem en annen gjennomslagskraft. Islamofobi på nettet er nært knyttet til tre trekk ved sosiale medier: globale nettverk, spredning av synspunktene av bestemte aktører (særlig «net troll» og «bots») og falske nyheter (Evolvi, 2018, s. 9). Evolvi (2019) beskriver hvordan en medieplattform som Twitter forsterker folks meninger, fordi man primært kommuniserer med andre som tenker likt som en selv (homophily). Hun mener at et sentralt trekk ved Twitter er at plattformen muliggjør raske, følelsesbaserte koblinger til andre – og i så måte skaper brukerne «affective news streams» (Evolvi, 2019, s. 388). Det er også slik at man på mange sosiale medier kan opptre anonymt, og anonymiteten senker trolig terskelen for å ytre seg hatefullt fordi man ikke blir holdt ansvarlig for sine ytringer (Dechesne et al., 2020).

Store aktører innen sosiale medier, som Facebook og Twitter, har de senere årene utviklet retningslinjer for hvilke typer ytringer de tillater (Facebooks såkalte «community standards»⁸ og Twitters «rules and policies»⁹). Facebook oppfordrer brukere til å rapportere poster som de anser for å være skadelige eller hatefulle, slik at Facebook kan vurdere om innholdet skal fjernes. Flere forskere påpeker at det er krevende for sosiale medieaktører å håndheve sine egne retningslinjer. Hatefulle ytringer blir ofte ikke fjernet, eller de blir stående lenge før de blir fjernet (se f.eks. Burke et al., 2020, s. 384; Oboler, 2016, s. 49). Én årsak er den massive mengden ytringer som til enhver tid blir lagt ut på Facebook. En annen årsak kan være at Facebook fokuserer på innholdet i et innlegg, og ser i mindre grad på praksiser knyttet til å spre slikt innhold gjennom likes, kommentarer og deling (Ben-David & Fernández, 2016, s. 1187).

3.6 Utløsende faktorer og sammenhengen mellom hendelser på nettet og utenfor nettet

En rekke studier tar for seg hvordan hendelser som skjer utenfor internettet (offline) kan fungere som utløsende faktorer (triggers) for hatefulle ytringer mot muslimer på internett. I flere av studiene finner forskere en kobling mellom terrorangrep i vestlige land, utført av radikale islamister, og hatefulle ytringer mot muslimer på internett i etterkant (Vidgen et al., 2021; Scharwächter & Müller, 2020; Evolvi, 2018; Awan & Zempi, 2016; Miller et al., 2016a; Miller et al., 2016b). Scharwächter og Müller (2020) fant for eksempel et systematisk forhold mellom terrorangrep i Europa og Nord-Amerika og hatefulle ytringer på Twitter. Anti-muslimske emneknagger (hashtags) blir ofte skapt i etterkant av terrorangrep (Evolvi, 2018, s. 1). Olteanu med kollegaer (2018) fant i sin studie at islamistisk vold førte til en økning i hatefulle ytringer mot muslimer på nettet, særlig ytringer som direkte oppfordret til vold mot muslimer. Forfatterne konkluderer med at «dessverre er bekymringene rundt en positiv feedback-loop mellom vold utenfor internett (offline) og hatefulle ytringer på nett (online) velfunderte»¹⁰ (Olteanu et al., 2018, s. 1).

Organisasjonen «Tell MAMA», som registrerer og analyserer angrep mot muslimer i Storbritannia, registrerte en tydelig økning i rapporter om islamofobe angrep både på og utenfor internett etter

⁸ Se https://www.facebook.com/communitystandards/objectionable_content for informasjon om Facebooks syn på hatefulle ytringer

⁹ <https://help.twitter.com/en/rules-and-policies#twitter-rules>

¹⁰ Vår oversettelse

terrorangrep i Storbritannia i 2017 (Tell MAMA, 2018). Mange av ytringene på nettet var ytterliggående. For eksempel tok en Facebook-bruker etter et terrorangrep i Manchester til orde for at alle muslimer skulle gasses, og en twitterbruker skrev at Anders Behring Breivik hadde «riktige ideer». Etter at en moské ble forsøkt påtent, var det mange som på Twitter og Facebook stilte seg positive til denne hendelsen og argumenterte for at alle moskeer bør ødelegges (Tell MAMA, 2018, s. 87). En annen bruker på Twitter skrev at «vi er i krig» og at man burde gå fra dør til dør og skyte alle muslimer (Tell MAMA, 2018, s. 89). Det kan også være hendelser i andre land som trigger hat mot muslimer lokalt (Awan & Zempi, 2016, s. 368). For eksempel førte skytingen i moskeen i Christchurch i New Zealand til en rekke angrep mot muslimer både på nett og ellers i Storbritannia, ifølge Tell MAMA. Mange av dem som angrep muslimer eller ytret seg hatefullt på nettet, refererte direkte til Christchurch-hendelsen (Tell MAMA, 2019, s. 52).

De fleste av undersøkelsene tar for seg sammenhengen mellom terrorangrep og hatefulle ytringer mot muslimer. Det er få studier som tar for seg andre hendelser. Ett unntak er en analyse av Facebook-sidene til to anti-muslimske bevegelser i Tsjekkia og Tyskland (Initiativ mot Islam og Pegida), hvor Hanzelka og Schmidt (2017) fant at hatefulle ytringer mot muslimer, migranter og flyktinger også ble trigget av andre typer hendelser enn terrorangrep utenfor nettet. I denne studien fant forskerne at triggere var knyttet til kriminalitet eller lovbrudd hvor innvandrere eller asylsøkere var involvert (f.eks. seksuelle overgrep mot kvinner i Tyskland; migranter som blokkerte jernbaneskiner på greske grenseoverganger; ulovlig grensekryssing), diskusjoner om asylpolitikk generelt (både nasjonalt og på EU-nivå, slik som kvotesystemet for flyktinger) eller «andre spesifikke handlinger», som kunne dreie seg om alt fra diskusjoner om ytringsfrihet og USAs militære aktivitet i Midtøsten, til det utenrikspolitiske forholdet til Tyrkia (Hanzelka & Schmidt, 2017, s. 154–155).

Islamistiske terrorangrep har altså en tendens til å trigge hatytringer mot muslimer på nettet, men slike angrep fører også til en økning i ytringer som imøtegår og kritiserer hatefulle ytringer mot muslimer (Olteanu et al., 2018). I en studie av responsen på Twitter etter terrorangrepene i Paris i 2015 finner Magdy og kollegaer (2015) at majoriteten av twittermeldinger som ble skrevet i timene etterpå, forsvarte muslimer – selv om det også var mange som anklaget muslimer for angrepene og knyttet islam til terror. Det var imidlertid stor variasjon i forholdet mellom positive og negative twittermeldinger avhengig av brukernes opprinnelsesland. Det var flest avsendere av twittermeldinger som forsvarte muslimer fra muslimske/arabiske land, som Saudi-Arabia, Jordan, Indonesia, Maldivene, Pakistan og Qatar. Bare to land hadde flere negative enn positive twittermeldinger, nemlig Israel og Nederland (Magdy et al., 2015, s. 6). Forskerne undersøkte også hvilke twittermeldinger som var mest populære, gjennom å se på hvor mye de ble retvitret. På tidspunktet da angrepet fant sted, var det valgkamp i USA. Blant de topp fem anti-islamske twittermeldingene som ble mest retvitret, var to fra president Trump; en var fra Ted Cruz, en annen presidentkandidat, som skrev en twittermelding som koblet islam til terror. I de to mest populære («populære» som i mest retvitrede) twittermeldingene som forsvarte muslimer, handlet budskapet om å skille mellom islam og den islamske staten ISIS (Magdy et al., 2015, s. 7).

En av studiene tar også for seg sammenhengen mellom hatefulle ytringer på Twitter og hatkriminalitet utenfor nettet i London, uavhengig av triggerhendelser (Williams et al., 2020). Forskerne fant at dersom det var en økning i hatefulle ytringer på nettet, rettet mot rase og religion, som ble sendt fra et område av byen, så var det samtidig en økning i hatkriminalitet utenfor nettet i samme område, rettet mot de samme grunnlagene. Det er ikke mulig å etablere en kausalsammenheng i datasettet som blir brukt, og forskerne peker på at sosiale medier antagelig bare er en av flere faktorer som påvirker hatkriminalitet utenfor nettet. For eksempel vil et nabolags særegenheter, slik som etnisk sammensetning og arbeidsledighetsrate, kunne ha betydning (Williams et al., 2020, s. 112). Studien kan likevel tyde på at sosiale mediers rolle i hatkriminalitet i

den «virkelige» verden er viktig. På bakgrunn av funnene argumenterer forfatterne for at hatkriminalitet bør ses som en prosess og ikke som enkeltstående hendelser, hvor det er en sammenheng mellom hatefulle ytringer i sosiale medier og hatkriminalitet utenfor nettet.

Tilsvarende opplever muslimer som blir utsatt for hatefulle ytringer og hatkriminalitet utenfor nettet disse typene hendelser som en sammenhengende del av det samme fenomenet og trusselbildet (Awan & Zempi, 2016; 2017). De som opplever hatefulle ytringer, lager ikke et tydelig skille mellom det som skjer på og utenfor nettet – de lever i frykt for at truslene som fremmes på nett, kan ramme dem i virkeligheten (se også Awan & Zempi, 2015). Hat mot muslimer blir ikke bare oppfattet som en trussel mot enkeltpersoner og deres familier, men blir også opplevd som en kollektiv trussel mot et større muslimsk fellesskap, som går på tvers av landegrenser. Det kan også anses for å være intensjonen bak hatefulle ytringer, som per definisjon er rettet mot grupper som helhet. Budskapet som blir oppfattet, er at muslimer ikke er velkomne og ikke hører til i vestlige samfunn (Awan & Zempi, 2016, s. 6).

3.7 Avsendere av hatefulle ytringer og spredning i nettverk

Det finnes relativt få studier som tar for seg dem som er *avsendere* av hatefulle ytringer mot muslimer på internett. Det er en hovedvekt av menn som skriver slike ytringer på nettet (Awan, 2016, s. 9; Hrdina, 2016, s. 42). Awan lager en typologi over hatytrere – for eksempel hvorvidt disse ytrerne primært produserte hatefulle ytringer eller delte ytringer, om de knyttet hets til oppkonstruerte hendelser, eller om de deltok i hets i etterkant av faktiske hendelser som et terrorangrep (Awan, 2016, s. 8). I en annen studie undersøker Awan såkalte nettroll, som han definerer som «folk som bruker sosiale nettverkssteder, som Twitter, for å produsere en hatkampanje mot muslimske samfunn» (Awan, 2014, s. 141).¹¹ Awan lager en typologi over åtte slike nettroll: 1) «tråleren», en person som aktivt leter etter Twitter-kontoer for å identifisere mål for hatefulle ytringer; 2) «lærlingen», en person som er relativt ny på Twitter, men som lar seg inspirere og læres opp av mer erfarne hatytrere; 3) «sprederen», en person som retvittr og sprer budskap med hatefulle ytringer; 4) «etterligneren», en person som bruker falsk profil eller bilder for å angripe personer; 5) «medhjelperen», en person som blir med i andres samtaler for å «ta» målgruppen; 6) «den reaktive», en person som retter hat mot muslimer etter bestemte hendelser, slik som et terrorangrep; 7) «bevegeren», en person som stadig skifter Twitter-profil for fortsette å angripe fra nye profiler; 8) «den profesjonelle», en person med svært mange følgere på Twitter, ofte med flere profiler, som målrettet angriper muslimer (Awan, 2014, s. 143).

I en annen studie studerte Hrdina (2016) hatytrere på Facebook i Tsjekkia i 2015, i tida da det kom mange migranter til Europa. Ytringene ble rettet mot flyktninger, migranter og muslimer, uten at hatytrerne skilte i særlig grad mellom gruppene; anti-islam-holdninger fyrte opp under anti-innvandringsholdninger og omvendt (Hrdina, 2016, s. 45). Den gjennomsnittlige hatytreren i studien var en mann mellom 35 og 50 år. Det var få studenter og eldre mennesker blant hatytrerne. Datamaterialet ga innsikt i utdanningen til hatytrerne bare for et lite utvalg (29 personer). Blant disse hadde nesten alle fullført videregående, og det var også en god del personer med høyere utdanning (primært innen IT, økonomi og ingeniørfag) i gruppen. De mest aktive hatytrerne hadde ikke flere Facebook-venner enn gjennomsnittet for utvalget (Hrdina, 2016, s. 42). De fleste produserte hatefulle ytringer relativt sjeldent, bortsett fra en liten gruppe med svært aktive hatytrere.

¹¹ Vår oversettelse

På grunn av dynamikken og teknologien på Facebook, hvor Facebook foreslår lignende saker for brukere som dem de har likt og kommentert på, skapes det raskt «samfunn» med hatytrere. Flesteparten av de hatefulle ytringene ble spredd i små og tilsynelatende private, men ikke lukka utvekslinger, slik som under artikler som blir lagt ut på individuelle profilsider. I utgangspunktet antok Hrdina (2016) at Facebook-sidene til høyreekstreme grupperinger var sentrale i å spre hatefulle ytringer, men dette så ut til å skje i mindre grad. Høyreekstreme grupper kunne ha rolle som opinionsskapere som satte agendaen, men spredningen av hatefulle ytringer så ut til å skje primært av uorganiserte, individuelle brukere som var motivert av sine egne holdninger (Hrdina, 2016, s. 43). Ekstreme grupperinger og politikere er altså ikke nødvendigvis den viktigste driveren bak spredning av hatefulle ytringer i sosiale medier.

De fleste hatefulle ytringer fant sted i tilknytning til diskusjoner om nyhetsartikler fra nyhetskanaler på nettet. Tilsynelatende nøytrale artikler kunne produsere mange hatytringer i kommentarfeltet og blant hatytrernes venner på Facebook, slik som artikler med informasjon om antall syriske flyktninger som ankom Europa. I den grad det ble lenket til andre sider, var det som regel ikke til alternative media, men til mainstream nyhetskanaler. Brukerne skapte sjelden innholdet som ble spredd selv, slik som artikler eller videoer (Hrdina, 2016, s. 43). Ifølge Hrdina var det bare en liten andel av ytringene i datasettet som inneholdt hatytringer som var så tydelig aggressive at de ville blitt fanget opp av algoritmer (Hrdina, 2016, s. 45).

I vårt litteratursøk fremkom det få studier som innebar at forskere faktisk intervjuet personer som har ytret seg hatefullt mot muslimer. Ett unntak er Herviks (2019) intervjuundersøkelse med danske personer som har kommet med islamofobe ytringer på nett. Hovedtrekkene er at de som har uttrykt hat mot muslimer på nett, ser på muslimsk kultur som inkompatibel med dansk kultur, at muslimer er farlige, og at den politiske venstresida er naiv, ikke forstår hvilken trussel islam utgjør, og unnskylder voldelige handlinger. Hervik (2019) skriver at kommunikasjonen rundt islam og muslimer blant innvandringskritikere som har ytret seg hatefullt, er preget av mye sarkasme, en likegyldighet til fakta og en forståelse av at muslimer og den «vestlige» befolkningen har inkompatible verdier og levemåter. Gjennom ytringene sine bidrar deltakere til å skape nettsamfunn (online «communities») hvor likesinnede finner tilhørighet og nye medlemmer rekrutteres.

Det er også relativt få studier av hvordan hatefulle ytringer sprer seg i sosiale nettverk. I en slik studie undersøkte Mathew med kollegaer (2019a) utbredelse og spredning av hatefulle ytringer på «Gab.com», en sosial nettverksside som er kjent for å ha mange brukere på ytre høyre-sida. Til tross for at Gab er kjent for å ha et større antall hatefulle ytringer enn for eksempel Twitter, fant forskerne at bare 1,32 prosent av alle ytringene kunne klassifiseres klart som hatefulle (Mathew et al., 2019a, s. 174). Personer som ytrer seg hatefullt, er tettere sammenkoblet i nettverk enn andre brukere. Forskerne fant at hatefulle ytringer sprer seg raskere og når et større nettverk enn ikke-hatefulle ytringer, og at innlegg fra hatefulle «ytrere» får dobbelt så mange svar og videresendinger (reposting) som ikke-hatefulle ytrere. Årsaken er imidlertid ikke at de som legger ut slike ytringer er mer populære enn andre, ut ifra antall følgere. Faktisk hadde ikke-hatefulle ytrere flere følgere – men hatefulle ytringer sprer seg likevel raskere, uavhengig av antallet følgere den som legger ut slike ytringer har. En årsak kan være at hatefulle ytrere ofte er svært aktive. Hatefulle brukere på Gab utgjorde bare 0,67 prosent av brukerne, men sto for hele 26,8 prosent av postene som blir lagt ut på nettstedet (Mathew et al., 2019a, s. 175).

3.8 Høyreekstremer partiers rolle i hatefulle ytringer mot muslimer på nettet

Flere studier undersøker hatefulle ytringer mot muslimer på nettsidene til høyreekstremer grupper. Et budskap i flere av studiene er at sosiale nettverksplattformer, gjennom islamkritiske budskap, bidrar til å føre ulike høyreorienterte nettverk tettere sammen og til å skape en transnasjonal, kollektiv identitet på høyresida (Gaudette et al., 2020, s. 1; Poole et al., 2019, s. 21; Cleland et al., 2018, s. 1153; Evolvi, 2018, s. 11).

Det er særlig tilstedeværelsen på internett blant høyreekstremer grupperinger i Storbritannia som har blitt undersøkt, slik som English Defense League og British National Party. I studien som tar for seg British National Party (Vidgen et al., 2021), benytter forskerne stordata fra Twitter for å utforske utbredelse av islamofobe ytringer, men også variasjon i atferd på nettet blant støttespillere til British National Party, som historisk har vært det mest sentrale høyreekstremer politiske partiet i Storbritannia. Forskerne fant at rundt fem prosent av twittermeldingene som ble sendt av partiets følgere på Twitter, var *eksplisitt* islamofobe. Omtrent dobbelt så mange av twittermeldingene, elleve prosent, var *implisitt* islamofobe, tidligere omtalt av de samme forskerne som «svak islamofobi», se punkt under («Skille mellom ulike former for islamofobi») (Vidgen et al., 2021, s. 8). Selv blant høyreradikale aktører er altså implisitt islamofobi betydelig mer vanlig enn eksplisitt islamofobe ytringer. Forskerne fant også stor variasjon i hvor aktive brukere var og hvorvidt aktiviteten deres begrenset seg til spesielle tidsvinduer, for eksempel i kjølvannet av et terrorangrep. Den store variasjonen, og de mange «passive» twitterbrukerne, tyder på at partiet har mange marginale støttespillere som ikke nødvendigvis kan betegnes som sterkt ideologisk overbevist. Forskerne fant at de fleste islamofobe ytringene ble fremsatt av et lite antall twitterbrukere, som var svært aktive.

I en annen studie sammenligner forskerne tekster fra den offisielle nettsida til English Defense League med tekster fra Facebook-sida deres (Cleland et al., 2018). Ifølge forskerne bygget brukerne av nettsidene opp en felles identitet rundt radikal opposisjon til islam og muslimer (Cleland et al., 2018, s. 1553). Målet til British Defense League er, ifølge organisasjonen selv, å bekjempe radikal islam. I tråd med dette fant forskere oftere ord knyttet til radikal islam på den offisielle nettsida til English Defense League enn på Facebook-sida deres, hvor også støttespillere deltok i diskusjonene. I disse diskusjonene var ikke søkelyset spesielt på radikal islam, men snarere på alle former for islam og muslimsk innvandring til Storbritannia.

I en studie av Facebook-sidene til høyreekstremer partier i Spania, fant forskerne på samme vis at partiene selv ikke la ut innlegg som inneholdt åpenlyse hatytringer, men direkte hatefulle ytringer ble ofte lagt ut av brukere i kommentarfeltene (Ben-David & Fernández, 2016). I denne studien undersøkte forskerne både tekst, bilder og lenker knyttet til Facebook-sidene til høyreekstremer partier i Spania. De fant at partier på ytre høyre for det meste delte lenker og bilder, snarere enn tekster (Ben-David & Fernández, 2016, s. 1177).¹² Ifølge forskerne brukte støttespillere til partiene på Facebook ofte strategier for å unngå at det de la ut, ble fjernet, slik som å legge ut en lenke til en side som inneholder hatefulle ytringer, uten å legge ut ytringene selv (Ben-David & Fernández, 2016, s. 1171).

¹² Se også Millar og kollegaer (2020) om rollen hyperlinks spiller i hatefulle ytringer

3.9 Innholdet i hatefulle ytringer

Det finnes en rekke studier som beskriver innholdet i hatefulle ytringer mot muslimer, ofte gjennom å lage typologier. Noen av studiene tar for seg ytringer på sosiale medier generelt, slik som på Twitter, mens andre undersøker ytringer på nettsteder som tydelig har en profil på ytre høyre-sida. I studier som tar for seg innholdet i ytringene, finner man mange slående likhetstrekk. En årsak kan som tidligere nevnt være at slike ytringer foregår på plattformer som per definisjon har en global karakter, hvor anti-islamske diskurser utvikler seg på tvers av landegrensene og skaper bånd mellom ulike høyreekstreme miljøer, i tillegg til å påvirke «mainstream-debatten» mer generelt.

Flere temaer går igjen på tvers av studiene. En sentral oppfatning er at muslimer utgjør en *demografisk trussel* i Europa, fordi det er et høyt antall innvandrere som er muslimer, og fordi muslimer får mange flere barn enn majoritets-europeere (Oboler, 2016; Ekman, 2015). Denne forestillingen er et kjent element i den såkalte «Eurabia»-teorien. Når muslimene blir mange nok, må europeiske politikere endre sin politikk for å tilfredsstille muslimske velgere. Gjennom dette demografiske skiftet og gjennom å få gjennomslag for islamsk lov og praksis på ulike samfunnsområder, infiltrerer muslimer vestlig politikk i det stille gjennom en «myk jihad» (Ekman, 2015, s. 1993). På sikt vil det føre til at muslimer innfører sharialov i Vesten, og potensielt borgerkrigliknende tilstander, fordi det kan være nødvendig med krig for å redde majoritetsbefolkningen fra islam (Cleland et al., 2018).

Tanken om at vestlige samfunn kommer til å bli dominert av islam og sharialov henger nært sammen med et annet sentralt tema, nemlig at islam er en trussel mot vestlig kultur og vestlige verdier (Evolvi, 2019; Cleland et al., 2018; Brindle, 2016; Oboler, 2016; Ekman, 2015). Islam blir fremstilt som en totalitær politisk ideologi som promoterer verdier som er diametralt motsatte av grunnleggende vestlige verdier (Ekman, 2015). Vestlig kultur blir fremstilt som liberal og progressiv, og islam blir fremstilt som illiberal og bakstreversk. Islam blir ansett for å være en grunnleggende kvinnefiendtlig religion som avviser vestlige normer knyttet til likestilling. Muslimske menns seksualitet blir beskrevet som sterk og ukontrollert, noe som fører til at de kan angripe ikke-muslimske kvinner. Det er hyppige referanser til voldtekt i islamofobe emneknagger (Evolvi, 2019). Mer generelt blir anti-muslimske holdninger legitimert ved at muslimer blir knyttet til undertrykkelse av kvinner, jøder og homofile (Brindle, 2016, s. 457).

Islams fremtidige dominans i Vesten blir muliggjort av den politiske *venstresida og liberale politikere*, som i praksis understøtter muslimenes plan om å få herredømme over Vesten (Gaudette et al., 2020; Oboler, 2016; Cleland et al., 2018; Brindle, 2016; Ekman, 2015). Det skjer gjennom at politikerne er naive og står for en liberal innvandringspolitikk og en multikulturell ideologi, som til syvende og sist undergraver vestlige verdier og er et ledd i underkastelsen for islam (Ekman, 2015, s. 1997). Venstresida unnskylder vold og terrorisme eller overser slike hendelser, og de favoriserer muslimske innvandrere fremfor trygghet og velvære for den ikke-muslimske majoriteten (Gaudette et al., 2020, s. 10–11). Det fremmes også et kritisk blikk på etablerte mediers representasjon av muslimer og majoritetsbefolkningen (Cleland et al., 2018, s. 1546).

Et annet dominerende fellestrekk på tvers av studiene er at muslimer blir beskrevet som en *sikkerhetstrussel*, og at islam blir beskrevet som en *grunnleggende voldelig religion* (Burke et al., 2020; Gaudette et al., 2020; Aguilera-Carnerero & Azeez, 2016; Awan, 2016; Brindle, 2016; Oboler, 2016; Ekman, 2015). Individuelle voldshandlinger utført av muslimer blir tolket som bevis for at islam er iboende voldelig. Islam knyttes til terror og vold, og det vises til sitater i koranen som legitimerer bruk av vold. I diskusjonene blir islamistisk ekstremisme og islam blandet sammen (Brindle, 2016; Burke et al., 2020). Moderate muslimer som tar avstand fra vold, beskyldes for å bidra til å skjule islams sanne natur. Muslimer blir mer generelt fremstilt som uærlige og manipulative (Oboler, 2016, s. 54). De blir knyttet til nyheter om voldtekt, vold mot ikke-muslimer,

seksuelt misbruk av barn og vold mot kvinner (Ekman, 2015). Awans (2016) studie av hatefulle ytringer mot muslimer på Facebook i etterkant av voldshandlinger utført av radikale islamister identifiserer liknende ytringer. Typisk innhold dreide seg om muslimer som terrorister; muslimer som voldtektsforbrytere; muslimske kvinner som en sikkerhetstrussel på grunn av klesplagg som dekker ansikt/hår og krig mellom muslimer. Aguilera-Carnerero og Azeez (2016) undersøkte innholdet i engelskspråklige twittermeldinger knyttet til emneknaggen «jihad». Twittermeldingene som benyttet emneknaggen formidlet ofte stereotypiske bilder av muslimer, som ble fremstilt som voldelige, ekstremistiske, som støttespillere av terrorisme mot Vesten, som tilbøyelige til å ta loven i egne hender, og som voldelige mot homofile, kvinner og personer fra andre religioner.

Mange av ytringene Oboler (2016) undersøkte, inneholdt trusler og oppfordring til vold mot muslimer – enten direkte eller indirekte. Flere av disse truende ytringene ble ikke fjernet, eller ble fjernet etter lang tid fra Facebook (Oboler, 2016, s. 49). Oboler (2016, s. 57) beskriver også at det har blitt opprettet falske sider på Facebook i muslimers navn, hvor det gis uttrykk for støtte til terrorisme og vold. Hensikten er å øke hat mot muslimer. Sistnevnte strategi har også blitt identifisert i en studie fra Danmark, hvor forskere gjennomførte en etnografisk studie av falske Facebook-sider som angivelig ble administrert av radikale islamister (Farkas et al., 2018a, 2018b). Sidene ble skapt av aktører fra ytre høyre for å styrke anti-muslimske følelser og holdninger som i sin tur skapte sterke reaksjoner mot innvandrere og muslimer generelt.

Ekman argumenterer for at en viktig funksjon av de hatefulle ytringene på Facebook er at de bidrar til å demonisere og dehumanisere muslimer. Muslimer fremstilles som djevelens håndlangere og som dyr, og kvinner i nikab sammenlignes med søppel. Slik dehumanisering fungerer som et ledd i å legitimere rasjonalen om at muslimer bør ekskluderes og fjernes. Ekman hevder også at de islamofobe diskursene han har identifisert på internett ikke bare må forstås som fenomener på den ytterliggående høyresida, men som tett sammenvevd med «mainstream» politiske diskurser, hvor det legges stor vekt på kulturforskjeller som essensialiserer kultur. Ekman argumenterer for at diskursen som florerer på islamofobe nettsider bidrar til å fyre opp under rasisme og negative holdninger mot muslimer, flytte debatten om islam lenger til høyre og skape større spillerom for krefter som vil begrense innvandring. Ekman knytter islamofobe diskurser til faktiske angrep mot muslimer og partier på venstresida i Europa, slik som angrep mot flere moskeer i Sverige og terroren rettet mot Arbeiderpartiet, utført av Anders Behring Breivik, i Norge. Breivik refererer til to av nettstedene Ekman undersøker, Jihad Watch og Gates of Vienna, i sitt manifest (Ekman, 2015, s. 1988–1999).

I en kvalitativ analyse undersøkte forskere twittermeldinger i Storbritannia i kjølvannet av to viktige hendelser i 2016: terrorangrepet i Brussel og beslutningen om Brexit (Miller et al., 2016a). De ulike dimensjonene av islamofobi som forskerne identifiserte, gjenspeiler mange av de samme temaene som er tatt opp i studiene over. Den ene dimensjonen dreier seg om islam som fiende, hvor islam blir ansett for å være iboende voldelig og fundamentalt fiendtlig innstilt til Vesten. En annen vanlig dimensjon er at muslimer er terrorister eller stiller seg støttende til terrorhandlinger. Videre inneholder mange twittermeldinger det som kan beskrives som «anger and abuse», det vil si sinte og nedsettende anti-islamske kommentarer. En tredje dimensjon er ønsket om å organisere handlinger utenfor nettet mot muslimer, noe som var særlig fremtredende etter terrorangrepene i Brussel. Et siste viktig tema i twittermeldingene er ideen om at mange muslimer er pedofile og dermed mer sannsynlig vil begå seksuelle overgrep mot barn.

Sistnevnte tema, nemlig koblingen mellom muslimer og pedofili, er et eksempel på en diskurs som gjør seg gjeldende i Storbritannia, men mindre i andre land. I et prosjekt som ble beskrevet tidligere, sammenligner Laurent (2020) hatytringer i Storbritannia, Italia og Frankrike. De finner mange fellestrekk på tvers av landene, men den komparative analysen viser også at det er noen forskjeller. I Storbritannia er islam og pedofili et gjennomgående tema, noe som antagelig er et resultat av

oppmerksomhet rundt pedofile nettverk blant personer med pakistansk bakgrunn. I Frankrike fremkommer pedofilitemaet i liten grad; her fokuseres det mer på bruk av hijab i offentligheten, som står i et spenningsfylt forhold til en sterkt sekulær kultur. Variasjonen i innholdet i hatytringene er ifølge forskerne et resultat av historien og kulturen i de bestemte landene og hvor de viktigste innvandringsgruppene kommer fra (Laurent, 2020, s. 2152).

I en australsk studie foretok Miller (2017) en automatisk tekstanalyse av innlegg på de to mest populære anti-islam-gruppene på sosiale medier i Australia. Han utforsker flere hypoteser om innholdet i ytringene ut ifra litteraturen om islamofobi. Hovedfunnene er at flest innlegg er knyttet til bekymringer for terrorisme og oppfatningen av islam som en politisk trussel mot demokratiet. I motsetning til flere av de andre studiene nevnt over, finner Miller liten støtte for hypotesen om at anti-islam-holdninger i Australia er knyttet til ideen om grunnleggende verdikonflikt, slik som oppfatningen av at muslimer har illiberale holdninger til kvinners, barns og homofiles rettigheter. Det er heller ikke særlig utbredt med bekymringer knyttet til religiøse praksiser. Forfatteren mener at dette gjør at Australia ligner på USA, men skiller seg noe fra Europa, hvor anti-islam-diskurser ifølge forskerne ofte har handlet mer om frykt for verdikonflikter og hvilke konsekvenser disse har for integrering.

I en studie ønsket forskerne å undersøke om innholdet i ytringene er ulikt avhengig av om de er rettet mot individer («directed hate speech») eller grupper («generalized hate speech») (EISherief et al., 2018). De fant at et trekk ved generaliserte hatytringer er at de oftest er rettet mot religioner, i motsetning til andre kategorier, slik som seksuell legning, kjønn eller nasjonalitet. Et annet trekk ved generaliserte ytringer er at de ofte inneholder ord knyttet til død, slik som utrydde, mord, halshugging og drap. I tillegg blir det i slike ytringer ofte brukt ord som benevner mengde, slik som «millioner» og «alle», og religiøse ord, slik som muslimer, jøder og kristne. Ordet «de» blir oftere brukt i generaliserte ytringer enn «vi». Mårettede ytringer («directed») er på sin side mer uformelle, personlige, sinte og mindre analytiske enn de generaliserte.

3.10 Skille mellom ulike former for islamofobi

I et par studier skiller forskerne mellom «svak» og «sterk» islamofobi (Vidgen & Yasseri, 2020), eller «liberal» og «illiberal» islamofobi (Evolvi, 2018). Svak islamofobi knyttes til ytringer som mer implisitt uttrykker negative holdninger mot muslimer, som at det er store kulturforskjeller mellom muslimer og andre medlemmer i samfunnet, og at muslimske kulturelle praksiser er fremmede og underlige. En annen dimensjon av svak islamofobi kan være at det understrekes at en person er muslim, selv om det ikke er relevant, slik at man implisitt knytter en problematisk hendelse (som pedofili eller vold) til muslimer generelt (Vidgen & Yasseri, 2020, s. 70). Distinksjonen mellom «liberal» og «illiberal» islamofobi er hentet fra Mondon og Winter (2017). Illiberal islamofobi er en åpen form for islamofobi som retter seg mot alle muslimer på grunn av deres religiøse tilhørighet. Liberal islamofobi, på den andre siden, er forankret i en forestilling om at muslimers kultur og praksiser bryter med grunnleggende, liberale verdier i vestlige samfunn, slik som kvinnefrigjøring og ytringsfrihet. Denne typen islamofobi er mer subtil og hviler på et underliggende skille mellom «gode» muslimer som har tilpasset seg og akseptert vestlige verdier, og «dårlige» muslimer som ikke ønsker å tilpasse seg (Evolvi, 2018, s. 3).

Siapera (2019) problematiserer skillet mellom åpenbart rasistiske og hatefulle ytringer, fremmet av aktører på ytre høyre-sida, og mer «akseptable» former for rasisme i form av mikroagresjoner og hverdagsrasisme. I studien undersøker Siapera twittermeldinger og kommentarer på Facebook-sida til en av de største irske nettavisene i etterkant av en voldshandling som mange trodde var terror, men som viste seg ikke å være det. Mens sosiale medieplattformer forsøker å hindre

åpenbart hatefulle ytringer gjennom sine «terms of service» (Twitter) og «community standards» (Facebook), blir mer moderate former for rasisme ansett som akseptable fordi det ville bli oppfattet som problematisk av hensyn til ytringsfriheten å slå ned på dem (Siapera, 2019, s. 3). På bakgrunn av datamaterialet mener Siapera at det i praksis er liten forskjell mellom disse formene for rasisme, at de ofte opptrer sammen, og at begge typer får sirkulere relativt uhindret i sosiale medier – også de grove ytringene. Å ha ulovlige hatefulle ytringer i fokus fører til et vilkårlig skille mellom ytringer det blir slått ned på og ytringer som ikke blir slått ned på, mens de er en del av samme diskurs. Siapera kommer imidlertid ikke inn på hvilke implikasjoner disse funnene har for ytringsfriheten.

3.11 Ytringer som imøtegår hatefulle ytringer

Sosiale medieplattformer forsøker å bekjempe hatefulle ytringer gjennom å fjerne innholdet eller blokkere brukere. Medieplattformer kan imidlertid også lett få kritikk for at slike handlinger er i strid med ytringsfriheten. Et alternativ er å bekjempe hatefulle ytringer med ytringer som imøtegår hatefulle ytringer (counter hate speech eller counterspeech) (Mathew et al., 2019b, s. 369). Som et ledd i denne strategien, har for eksempel Facebook satt i gang egne prosjekter for å øke engasjementet rundt motnarrativer.¹³ I «hatemeter»-prosjektet (Laurent, 2020), som kontinuerlig samler inn data om hatefulle ytringer mot muslimer på nettet, har forskerne i samarbeid med frivillige organisasjoner laget en motnarrativ-funksjon (counter-narrative-function) som fungerer slik at man skriver inn en hateful ytring i et program, som deretter presenterer en liste med fire ulike responser som man kan bruke eller redigere ved behov. De redigerte versjonene blir også lagret i databasen. Denne funksjonen skal bidra til å forebygge hatefulle ytringer. Responsene er skrevet av ansatte i frivillige organisasjoner og er modellert etter typiske hatytringer. Nye anti-hatytringer blir automatisk klassifisert sammen med de opprinnelige ytringene, slik at det er sannsynlig at responsen passer. Funksjonen skal gjøre det enklere og raskere å imøtegå hatefulle ytringer (Laurent, 2020, s. 2148–2149).

I en studie av ytringer som imøtegår hatefulle ytringer (counterspeech) på YouTube (Mathew et al., 2019b), ble over 13 000 ytringer manuelt klassifisert i ytringer som imøtegår (counterspeech) eller ytringer som ikke imøtegår hatefulle ytringer (non-counterspeech), altså ytringer som viser enighet med den opprinnelige, hatefulle ytringen eller nye hatefulle ytringer. Datasettet ble senere brukt til å trene opp en algoritme som klassifiserte ytringer som imøtegår hatefulle ytringer (counterspeech) og ytringer som ikke imøtegår hatefulle ytringer (non-counterspeech) automatisk. Deretter ble motytringene klassifisert i ulike strategier, slik som for eksempel bruk av humor, bruk av fakta, advarsel om hvilke konsekvenser hatprat kan få, og fiendtlig språkbruk (Mathew et al., 2019b, s. 371–373). Forskerne fant at den minst effektive strategien var å benytte fiendtlig språk. Motytringer hvor fiendtlig språk ble benyttet, møtte mer motbør, også blant brukere som i utgangspunktet ønsket å vise sin støtte til grupper som ble utsatt for hatefulle ytringer (Mathew et al., 2019b, s. 379).

Forskerne fremhever mulighetene som ligger i sosiale medier til å bekjempe netthets og falske nyheter. De fant at kommentarer som inneholdt ytringer som imøtegikk hatefulle ytringer fikk mange flere likes enn ytringer som ikke imøtegikk hatefulle ytringer (non-counterspeech comments) (Mathew et al., 2019b, s. 370) – noe som er interessant, med tanke på andre studier som viser at hatefulle ytringer sprer seg raskere og til et større nettverk enn andre typer ytringer. Downing og Dron (2020) gjorde lignende observasjoner i en studie som ser på twittermeldinger i etterkant av en brann i England (Grenfell) hvor mange muslimer var berørt og også bidro til å redde mennesker fra det brennende bygget. De fant at det var dobbelt så mange twittermeldinger som imøtegikk

¹³ <https://counterspeech.fb.com/en/#>

«fake news» som fremstilte muslimer i et dårlig lys, som dem som spredte slike narrativer. I dette tilfellet var Twitter en plattform som bidro til å spre positive narrativer om muslimer og som muliggjorde at folk kunne imøtegå falske nyheter og hatefulle meldinger. Slike motytringer kan være enklere på Twitter enn på Facebook, fordi kommunikasjonen på Twitter er åpen for alle, mens mye kommunikasjon foregår i mer lukkede rom på Facebook.

Poole og kollegaer (2019; 2020) har i et prosjekt undersøkt potensialet for motnarrativer (counternarratives) som imøtegår hatefulle ytringer knyttet til muslimer og islam på Twitter. De undersøkte twittermeldinger knyttet til emneknaggen #stopislam, som vokste etter terrorangrepene i Brussel i 2016 utført av IS. Emneknaggen trendet på Twitter, ikke bare fordi den ble brukt for å spre hatefulle ytringer mot muslimer, men også fordi mange brukte emneknaggen for å ta til motmæle mot slike ytringer. I en av artiklene (Poole et al., 2019) skriver forskerne at de opprinnelig hadde tenkt å fokusere på motnarrativer (counternarratives), men fant fort ut at disse henger så tett sammen med de opprinnelige hatnarrativene at ytringene må undersøkes i sammenheng. Motnarrativet viser at dominerende hatnarrativer ikke nødvendigvis blir stående uten motstand, men en nettverksanalyse tyder på at nettverket mellom de som promoterer anti-muslimske twittermeldinger er tettere sammenvevd enn dem som promoterer motnarrativet, som er løsere forbundet og mer ad hoc (Poole et al., 2019, s. 28). Slik sett kan brukere som fremmer motnarrativer beskrives mer som et løst sammenkoblet fellesskap sammenlignet med grupperinger på høyresida på Twitter (Poole et al., 2019, s. 29). Når emneknaggen #stopislam ble brukt også for å fremme et motnarrativ, er det et eksempel på nettkapring (online hijacking) (Jackson & Foucault Welles, 2015), det vil si at ytrere på internett begynner å bruke en emneknagg for å fremme et helt annet budskap enn det som opprinnelig ble forbundet med emneknaggen. Slik kan en annen tolkning av en hendelse få gjennomslag i den diskursive nettvirkeligheten. En slik kapring av en emneknagg som opprinnelig promoterer et hatefullt budskap viser potensialet som ligger i sosiale medier når det gjelder å imøtegå hatefulle ytringer.

Twitter er i mindre grad et ekkokammer enn Facebook, hvor profilene ofte er lukket og nettverkene er mer selektive. På Twitter kommer folk med ulike meninger oftere i kontakt med hverandre (Poole et al., 2020, s. 11). Det var flere menn enn kvinner som benyttet emneknaggen #stopislam. De mest aktive aktørene hadde tilknytning til høyresida/ekstreme høyre, og mange identifiserte seg som kristne og konservative i egne Twitter-biografier (Poole et al., 2020, s. 10). De mest delte twittermeldingene under emneknaggen #stopislam, var twittermeldinger som forsvarte islam og muslimer, men *responsene* på disse twittermeldingene var primært preget av et anti-islamisk budskap (Poole et al., 2020, s. 8). Twittermeldinger som representerte et motnarrativ ble mest delt av muslimer, og av personer fra Storbritannia og land i Midtøsten og Nord-Afrika (Poole et al., 2019, s. 23–24). Store nyhetsorganisasjoner (Al Jazeera, CNN, The Washington Post osv.) var aktive i å promotere motnarrativet (Poole et al., 2019, s. 26). Det viktigste temaet i motnarrativet var å argumentere mot koblingen mellom islam og terror (Poole et al., 2019, s. 24). Twittermeldinger som brukte #stopislam for å fremme et motnarrativ ble mest delt det første døgnet etter terrorangrepet, og hadde i så måte kortere levetid enn twittermeldinger knyttet til emneknaggene som var islamfiendtlige – emneknaggen levde videre og var viktig også etter senere terrorangrep (Poole et al., 2019, s. 25).

Islamfiendtlige twittermeldinger under #stopislam ble oftest produsert og delt av brukere i USA. Da forskerne undersøkte relaterte emneknagger, fant de at mange av disse kunne knyttes til høyrekonservative miljøer i USA som blant annet promoterte Trumps presidentkampanje (Poole et al., 2019, s. 16). USA var også det eneste landet hvor anti-islam-twittermeldinger helt klart var i overvekt sammenlignet med twittermeldinger som imøtegikk disse. Andre land som hadde marginalt flere anti-islam-twittermeldinger, var Canada, Australia og Tyskland (Poole et al., 2019, s. 20). Anti-muslimske twittermeldinger ble også knyttet til en agenda mot den politiske venstresida

(Poole et al., 2019, s. 24). Resultatene bekrefter tidligere funn som tilsier at Twitter blir brukt av ytre høyre for å mobilisere støttespillere og spre budskapet sitt gjennom å skape nettverk blant konservative miljøer.

Poole og kollegaer (2020) er kritiske til forståelsen av at mer rom for «dialog» er den beste medisinen mot hatefulle ytringer – selv om det kan oppstå motnarrativer på Twitter, er båndene mellom høyreorienterte grupper såpass sterke at maktforholdet mellom det hatefulle narrative og motnarrativet er asymmetrisk, i favør av det hatefulle budskapet. Motnarrativet bidrar også i en viss forstand til å reprodusere det originale narrative, fordi det blir bygget på premissene til den opprinnelige, hatefulle ytringen. I stedet kan det være viktig å skape ytringer som imøtegår hatefulle ytringer (counter speech) i motsetning til motnarrativer, som skaper helt nye fortellinger (Poole et al., 2020, s. 23).

3.12 Stordatastudier

De store medieplattformene som Facebook og Twitter har, som tidligere nevnt, de siste årene utviklet såkalte «community guidelines» eller lignende, som beskriver hvilket ansvar medieplattformene har for å bekjempe hatefulle ytringer og hva slags ytringer som kan anses for å være såpass alvorlige at de burde fjernes. Identifisering av hatefulle ytringer i sosiale medier har så langt vært avhengig av at brukere rapporterer inn slike ytringer eller at medieplattformene overvåker ytringer manuelt, noe som er krevende, med tanke på den store mengden ytringer på sosiale medier. For raskere å kunne identifisere hatefulle ytringer og for å undersøke hvor utbredt slike ytringer er i sosiale medier, har forskere de senere årene benyttet automatiserte metoder for å innhente store mengder data fra medier som YouTube (f.eks. Ottoni et al., 2018), Twitter (f.eks. Vidgen et al., 2021; Lingardi et al., 2020; ElSherief et al., 2018; Poletto et al., 2017; Burnap & Williams, 2015), Reddit (f.eks. Olteanu et al., 2018) og Facebook (f.eks. Hrdina, 2016).

En fremgangsmåte er å lage et korpus med ord på forhånd som assosieres med bestemte grupper eller temaer, som deretter benyttes som søkeord («lexicon-based approach») for å fange opp ytringer som kan være relevante (se f.eks. Lingardi et al., 2020). I noen studier blir det deretter foretatt et tilfeldig utvalg av ytringer som kategoriseres og analyseres nærmere gjennom innholdsanalyse/manuelle metoder (se f.eks. Hrdina, 2016). I andre studier benyttes det maskinlæringsmetoder for automatisert klassifisering av hatefulle ytringer i stedet. Slike maskinlæringsmetoder er enten «supervised» eller «unsupervised» (Nadim et al., 2018, s. 47). Metoder som er «unsupervised» innebærer ikke manuell klassifisering i første runde (se f.eks. Zhang & Luo, 2019; Miller, 2017;). Med «supervised» maskinlæringsmetode er manuell koding et viktig første steg i klassifiseringsprosessen, men slik manuell koding er ikke et mål i seg selv, bare et første trinn i maskinlæringen (se f.eks. Vidgen et al., 2021). Et utvalg ytringer klassifiseres manuelt for å lage et såkalt «treningssett», hvor to eller flere personer koder samme ytringer for å sikre at det er noenlunde enighet om hvordan ytringene skal klassifiseres. I noen studier benyttes det såkalte «crowdsourcing-tjenester» på nettet, hvor forskere kan finne personer til å klassifisere data manuelt ut ifra de kriteriene de ønsker (ett eksempel er «Crowdfower», se Magdy et al., 2015). Etter at ytringer har blitt klassifisert manuelt, brukes disse klassifiseringene til å «lære opp» en algoritme, som deretter kan gjennomføre klassifisering av et stort datamateriale.¹⁴ Det er behov for å kode mellom 2000 og 3000 eksempler på hatefulle og ikke-hatefulle ytringer for å lage treningssettet (Nadim et al., 2018, s. 48).

¹⁴ Se rapport fra Institutt for samfunnsforskning for et overblikk over bruk av maskinlæring og andre metoder i omfangsstudier av hatefulle ytringer (Nadim et al., 2018)

Det har blitt utviklet dataplattformer som bidrar til å samle inn, lagre og analysere stordata om hatefulle ytringer mot muslimer på nettet. Ett eksempel er «the Hatemeter platform»¹⁵, som er utviklet av forskere fra Frankrike, Storbritannia og Italia (Laurent, 2020). Data samles inn på fransk, engelsk og italiensk, men metoden kan også brukes på andre språk. Plattformen brukes til å samle inn data daglig fra Twitter og YouTube ved hjelp av bestemte søkeord. I prosjektet overvåker man emneknagg-trender, men også det man omtaler som «hatefulle ytrere» («hate speakers») for å identifisere nettverk av hatytrere og hvordan de samhandler. Dataene brukes også til å produsere såkalte motnarrativer (counternarratives), det vil si ytringer som imøtegår islamofobe budskap i sosiale medier (se avsnitt om dette over). Et annet eksempel er «Hatebase»¹⁶, en nettressurs som også bruker algoritmer for å overvåke og analysere hatefulle ytringer. Databasen har en liste med ord og ytringer på over 95 språk som regnes som hatefulle (se EISherief et al., 2018 for et eksempel på en studie som har benyttet Hatebase).

3.12.1 Begrensningen med stordatastudier og utfordringen med klassifisering

Som allerede nevnt, er en begrensning med stordatastudier så langt at de bare fanger opp tekst, men ikke visuelt innhold, slik som bilder og symboler. En annen begrensning ved bruk av automatiserte datainnsamlingsmetoder er at det vil være en del ytringer som er hatefulle som ikke fanges opp. For eksempel vil ytringer som benytter ord og uttrykk som faller utenfor forhåndsdefinerte søkeord, falle utenfor. I sin studie viser Hrdina (2016) for eksempel til at en uttalelse som «vi må utrydde disse menneskene» ikke ville blitt inkludert i datasettet ut ifra hans søkekriterier, mens samme uttalelse ville blitt fanget opp dersom den inkluderte «syrere» eller «afrikanere», som er lagt inn som søkeord (Hrdina, 2016, s. 40). Andre forskere påpeker også at det er vanlig at målgruppen blir utelatt i setningen, og i stedet blir erstattet av et nedsettende uttrykk, som «pakk» eller «skitt». Målgruppen blir først tydelig når man ser ytringen i kontekst (Baumgarten et al., 2019, s. 11). Man kan anta at slike ytringer er vanskelige å fange opp for en algoritme. En konsekvens kan være at studier som benytter automatiserte datainnsamlingsmetoder til en viss grad underestimerer hvor mange hatefulle ytringer som finnes (Hrdina, 2016, s. 40).

Dersom man bruker maskinlæringsmetode for å klassifisere ytringer i materialet som har blitt innhentet, er det også begrensninger med tanke på hvor finmaskede klassifiseringene kan være. Som regel klassifiseres ytringene i slike studier relativt enkelt, i to kategorier: hatefulle eller ikke hatefulle. I noen tilfeller har forskerne forsøkt å fange opp ytterligere nyanser, slik som å skille mellom «sterk» og «svak» islamofobi (Vidgen & Yasseri, 2020). Forskerne erfarte imidlertid at mens algoritmen var relativt god til å skille mellom ikke-islamofobe og sterkt islamofobe ytringer, greide den ikke å skille like godt mellom twittermeldinger som var sterkt eller svakt islamofobe og mellom svake og ikke-islamofobe ytringer (Vidgen & Yasseri, 2020, s. 74). Kategorien «svak islamofobi» var altså vanskelig å fange opp gjennom automatisert klassifisering. Man kan imidlertid tenke seg at en algoritme vil kunne skille, for eksempel, mellom hatefulle ytringer som «bare» er nedsettende og ytringer som fremsetter konkrete trusler om vold. Jo flere kategorier eller dimensjoner man ønsker å fange opp, jo større datasett er det behov for, med tilsvarende økt arbeid for å klassifisere ytringene som skal brukes som treningssett (dersom det er snakk om «supervised» maskinlæring).

Dersom man ønsker å lage flere og mer nyanserte kategorier, kan det ut ifra litteraturen se ut til å være en bedre løsning å klassifisere et tilfeldig utvalg av ytringer manuelt i stedet for å lære opp en algoritme til å foreta klassifiseringen. Om man gjør det, kan man klassifisere ytringer etter en rekke ulike kriterier, slik LDO (2018) gjorde i sin studie, hvor de blant annet undersøkte hvilket

¹⁵ <http://hatemeter.eu/>

¹⁶ <https://hatebase.org>

grunnlag ytringen rettet seg mot, om ytringen inneholdt skjellsord, og hvorvidt ytringen fremstår stigmatiserende, nedsettende, truende osv.

En annen utfordring er at mange ytrere unngår å bruke direkte hatefulle ord og vendinger som de vet er ulovlige. Ytringene er ofte subtile, og det hatefulle budskapet fremkommer tydelig først når ytringen leses i kontekst (Laurent, 2020, s. 2151). Det er særlig vanskelig å fange opp ironi og sarkasme, som ofte benyttes i hatefulle ytringer (Vidgen & Yasseri, 2020, s. 74; Baumgarten et al., 2019, s. 10-11). Når hatefulle ytringer fremsettes med en humoristisk tone, kan det være en defensiv taktikk, fordi de som har fremsatt ytringen kan si at det bare var humoristisk ment dersom de får kritikk. Med ironi kan ytreren gi uttrykk for meninger uten å direkte artikulere dem (Hervik, 2019). I et prosjekt om tysk- og danskspråklige hatefulle ytringer på Facebook og Twitter fant forskerne at mange hatefulle ytringer begynte med ironiske språklige konstruksjoner som «de å-så-fredelige» pluss målgruppen, etterfulgt av negative ytringer. «Å-så ...» ble også brukt i motytringer, som i «de å-så-onde muslimene», for å vise avstandtaking til slike holdninger (Baumgarten et al., 2019, s. 11).

Kommentarer på nettet er ofte preget av feilstavinger og forkortelser som ikke blir fanget opp av søkeord. Mange ord kan brukes både i negative og positive sammenhenger (EISherief et al., 2018). Når det gjelder ulovlige ytringer, er det vanskelig for politiet å gå videre med tips om hatefulle ytringer på nettet der tolkningene er flertydige, selv om også slike ytringer har blitt straffeforfulgt i senere tid. Loven forutsetter at budskapet ikke er til å misforstå. Emneknagger kan være nyttige for å identifisere om ytringer er hatefulle, fordi emneknagger ofte indikerer ideologisk tilhørighet. Samtidig brukes emneknagger også av personer som imøtegår hatefulle ytringer, og de kan også brukes ironisk, både av hatytrere og andre. Av disse årsakene vil kvalitative analyser ofte være et viktig supplement til stordataanalysene (Laurent, 2020, s. 2151).

3.13 Oppsummering

Det finnes begrenset med nordisk forskning om hatefulle ytringer på internett mot muslimer. Flesteparten av studiene i kunnskapsoppsummeringen er fra andre land.

Litteraturgjennomgangen viser at det ikke er en omforent definisjon av hatefulle ytringer i studiene, men de fleste forskerne legger til grunn en bred definisjon av hatefulle ytringer, som inkluderer både lovlige og straffbare ytringer. Innholdet i hatefulle ytringer har ikke nødvendigvis endret seg så mye over tid, men de sprer seg raskere enn før gjennom sosiale medier, hvor folk kommuniserer med likesinnede og kan ytre seg anonymt. Sosiale nettverksplattformer bidrar til å føre ulike høyreorienterte nettverk tettere sammen og til å skape transnasjonale, kollektive identiteter på høyresida. Forskingen viser at det er en relativt lav andel ytringer på sosiale medier som er eksplisitt hatefulle, og at slike ytringer gjerne blir produsert av et lite antall, svært aktive brukere. Noen studier tyder imidlertid på at hatefulle ytringer sprer seg raskere enn andre typer ytringer. Hatefulle ytringer på nett blir gjerne trigget av ekstreme hendelser, som terrorangrep.

I den internasjonale litteraturen er det relativt få diskusjoner om avveiningene mellom å bekjempe hatefulle ytringer, og ytringsfrihet. Slike diskusjoner har vært sentrale i politiske debatter i Norge. En måte å bekjempe hatefulle ytringer på, uten nødvendigvis å fjerne ytringene, er å imøtegå dem. I noen prosjekter fokuserer forskere og frivillige organisasjoner på å lage databaser med såkalte «motnarrativer» som kan benyttes i møte med hatefulle ytringer på nettet.

Flere kvalitative studier tar for seg innholdet i ytringene, gjerne gjennom å skape typologier. Trekk som går igjen på tvers av disse studiene er en forestilling om at muslimer er i ferd med å ta over Europa, kritikk av den politiske venstresidas manglende vilje til å ta inn over seg trusselen fra islam,

fremstillingen av muslimer som iboende voldelige og farlige, en tett kobling mellom muslimer generelt og terrorisme, og fremstilling av kulturelle og verdimeslige forskjeller mellom «islam» og «Vesten» som grunnleggende og uoverstigelige. I tillegg til de mange fellestrekkene, er det også noen dimensjoner som er mer vanlige i enkelte land, slik som koblingen mellom islam og pedofili i Storbritannia.

Flere av de andre studiene vi har gjennomgått her, er såkalte stordatastudier, hvor det benyttes automatiserte metoder for å samle inn store mengder data fra sosiale medier. I tillegg benyttes ofte maskinlæringsmetoder for å klassifisere ytringene. Slike studier er nyttige fordi de kan fortelle noe om omfanget av hatefulle ytringer, og fordi de på sikt kan bidra til effektiv og rask identifisering av hatefulle ytringer på sosiale plattformer, noe som er krevende å gjøre dersom man beror på menneskelig rapportering og overvåking. Noen av utfordringene er at algoritmene ikke fanger opp innhold med bilder, symboler, o.l., at det er vanskelig å gjøre mer nyanserte klassifiseringer, og at bruk av ironi og sarkasme gjør at hatefulle ytringer kan være vanskelige å fange opp.

Det er færre studier som undersøker nettverk og spredning av hatefulle ytringer. Det er også relativt få studier som tar for seg avsendere av hatefulle ytringer eller hvordan hatefulle ytringer på internett påvirker muslimer.

I denne kunnskapsoppsummeringen har vi tatt for oss forskning om hatefulle ytringer mot muslimer på internett. I søket brukte vi en rekke ulike ord for å finne frem til relevant litteratur, men mye av litteraturen som blir beskrevet her, dreier seg om ytringer og holdninger i mer ytterliggående miljøer. Det er sannsynlig at forskning som tar for seg negative holdninger mot muslimer i befolkningen mer generelt, i liten grad har blitt fanget opp i vårt søk. Forskning tyder på at moderate og mer ekstreme ytringer henger nært sammen, og således kan anses for å være en del av samme fenomen. I en fremtidig kunnskapsoppsummering ville det være nyttig å undersøke hvordan skepsis mot muslimer kommer til uttrykk i mindre ytterliggående debatter og miljøer.

4 Analyser av omfang – stordataanalyser

4.1 Metode

For å analysere omfanget av hatefulle eller diskriminerende ytringer om muslimer på norske sosiale medier har vi anvendt stordata – en analysemetode som innebærer innhenting og analyse av store mengder informasjon. Stordata er en samlebetegnelse for data som er så store at de krever mer enn vanlig datakraft for å samles inn, lagres og analyseres. Begrepet stordata brukes ofte ikke bare for å betegne selve dataene, men også for å beskrive de nye problemstillingene slike data reiser, både teknisk, juridisk og etisk (Nadim et al., 2018, s. 45).

I denne studien er våre data offentlige kommentarer på sosiale medier. Mer spesifikt har vi undersøkt tekstinnhold av kommentarer på Facebook og Twitter. Bilder eller memes, eventuell tekst inne i bildene eller memene, er ikke innhentet av våre kvantitative metoder. Selv om vi kun har innhentet offentlige kommentarer, er det ikke gitt at individer har full forståelse for at det de skriver på sosiale medier er tilgjengelig for alle og kan bli brukt som data av forskere. Derfor har vi i prosjektet begrenset typen data vi har innhentet for å sikre folks anonymitet. Prosjektet har en personvernkonsekvensvurdering (DPIA) som er godkjent av både OsloMet og Norsk senter for forskningsdata (NSD).

Å innhente data fra Twitter er ganske enkelt via Twitters egen API (Application Programming Interface). Facebook, derimot, er svært lukket etter Cambridge Analytica-skandalen, og tilbyr ingen

enkel løsning slik Twitter gjør. Vi utviklet derfor en egen løsning for å automatisere innhenting av kommentarer fra Facebook.

Klassifiseringsmodellen vi har utviklet i dette prosjektet involverer mange parametere som gjør det vanskelig å se hvordan vi kom frem til resultatene. Dette er det såkalte svart-boks-dilemmaet som preger mange av maskinlæringsmetodene (Rudin, 2019), og er knyttet til åpenhet og transparens. Vi har prøvd å begrense dette problemet ved å gjøre alle kodinger for innhenting av data, utvikling av modeller og analyser tilgjengelige i GitHub.¹⁷ I tillegg har vi skrevet kodene enten på R eller Python, som er de mest etablerte programmeringsspråkene for stordataanalyser og maskinlæring.

4.1.1 Maskinlæring

I denne studien har vi brukt maskinlæring, som er en spesialisering innen kunstig intelligens der statistiske metoder brukes for å få datamaskiner til å finne mønstre i store datamengder. Vi har utviklet en statistisk modell¹⁸ og trent denne til å gjenkjenne og finne hatefulle ytringer, altså å skille hatefulle ytringer fra ikke-hatefulle ytringer. Den formen for maskinlæring kalles veiledet læring. I veiledet læring finner den statistiske modellen et mønster ved hjelp av mange eksempler – et såkalt annotert datasett. I dette tilfellet har modellen fått eksempler på både hatefulle og ikke-hatefulle ytringer for å lære hva det er som kjennetegner begge typer ytringer. Det å lære opp en modell kalles også å trene opp en klassifiseringsmodell. Klassifiseringsmodellen kan søke gjennom store mengder data og skille mellom hatefulle og ikke-hatefulle ytringer.

Ved hjelp av klassifiseringsmodellen vi har trent opp, har vi anslått omfanget av hatefulle ytringer mot muslimer på Twitter og Facebook. Å trene opp en slik modell består av tre faser:

1. Å definere hatefulle ytringer
2. Å lage et annotert datasett
3. Valg av algoritme, testing og forbedring av modellens ytelse

4.1.2 Å definere hatefulle ytringer

Vi har definert hatefulle ytringer mot muslimer ved hjelp av trekk som går igjen i kunnskapsoppsummeringen vi har gjort, og samtaler med referansegruppen. I definisjonen av hatefulle ytringer inkluderte vi kommentarer med referanser til at muslimer

- er kollektivt ansvarlige for vold, terror og politikk i andre land
- er iboende voldelige og farlige og at det er en tett kobling mellom muslimer generelt og terrorisme
- arbeider for en islamisering eller kulturell dominans av Norge og/eller Europa, gjerne i samarbeid med politiske partier og medier (konspirasjonsteorien om Eurabia)
- er i ferd med å overta Europa
- har en særegen mentalitet som er uforenlig med vestlige verdier og livsstil
- ødelegger Norge, og muslimske plagg og symboler bør forbys.
- bør sendes ut av landet og forbys adgang

¹⁷ <https://github.com/upretiramesh/OsloMet-KARTLEGGING-AV-OMFANGET-AV-HATEFULLE-ELLER-DISKRIMINERENDE-YTRINGER>

¹⁸ Empiriske/statistiske modeller er grunnleggende empiriske, men gjør i tillegg bruk av sannsynlighetsteori. En statistisk modell av et system kjennetegnes ved flere parametere som beskriver relasjoner mellom variabler med verdier fra gitte fordelinger. Statistisk modellering har vanligvis som mål å estimere parameterverdiene. Ved å observere og analysere data for et fenomen, kan en ved hjelp av statistiske metoder avdekke hvilke variabler som er relevante og hvilke av modellens sammenhenger som er de vesentligste (Helgesen et al., 2019).

- dehumaniseres

I datasettet har vi kun skilt mellom hatefulle og ikke-hatefulle ytringer. Vi har ikke vurdert graden av alvorlighet i ytringene. Kunnskapsoppsummeringen beskriver begrensninger ved bruk av stordata.

4.1.3 Annotert datasett

Klassifiseringsmodellen trenger mange eksempler for å lære å skille hatefulle ytringer fra ikke-hatefulle ytringer. Vi laget et annotert datasett bestående av hatefulle og ikke-hatefulle ytringer for å trene klassifiseringsmodellen. Antall hatefulle ytringer er lavt sammenlignet med ikke-hatefulle ytringer, slik også andre studier av hatefulle ytringer viser (Vidgen & Yasseri, 2018). Det annoterte datasettet må dermed være stort for å ha tilstrekkelig antall eksempler av hatefulle ytringer. Å finne nok eksempler på hatefulle ytringer til å lage et annotert datasett man kan trene modellen med, er tidkrevende. Vi har brukt forskjellige tilnærminger for å identifisere hatefulle og ikke-hatefulle ytringer som kunne inngå i det annoterte datasettet. På Facebook har vi samlet kommentarer som omhandlet muslimer eller islam under nyhetssaker delt av etablerte og alternative nyhetsredaksjoner på Facebook-sidene deres. Vi antok at såkalte sinnafjes kunne øke sannsynligheten for å finne hatefulle kommentarer, og derfor samlet vi kommentarer fra nyhetssaker som hadde et høyt antall sinnafjes. På Twitter, som har relativt få hatefulle ytringer mot muslimer, har vi lett på personlige profiler og gruppeprofiler som er kjent for å være negative overfor muslimer og islam. På disse Twitter-profilene var det få kommentarer, men mange kunne annoteres som hatefulle. Vi har også samlet kommentarer på SIAN sin egen nettside, og på noen diskusjonsgrupper på VG debatt-nettsida der temaet var muslimer eller islam. Kommentarene fra disse kildene utgjør det annoterte datasettet vårt, og består av 3277 kommentarer. Av disse er 2471 kommentarer manuelt merket som «ikke-hatefulle» og 806 er manuelt merket som «hatefulle». To av forskerne i prosjektet gjorde den manuelle merkingen av ytringene, den såkalte annoteringen av hatefulle og ikke-hatefulle ytringer. Først merket en av forskerne ytringene som hatefulle eller ikke-hatefulle. Deretter reviderte den andre forskeren merkingen. Til slutt fjernet vi alle ordene eller symbolene som ikke var relevante, for eksempel URL-er og såkalte «stoppord» (stop words) som artikler og preposisjoner.¹⁹

4.1.4 Valg av algoritme, testing og forbedring av ytelse

Vi prøvde forskjellige alternativer med hensyn til valg av algoritmer. Algoritmen som ga best resultat, var et kunstig nevralt nettverk på en vanlig CPU-prosessor. Kunstig nevralt nettverk er en type *dyp læring* som gir gode resultater på klassifiseringer av komplekse data som tekst (LeCun et al., 2015; Dvergsdal, 2019). Denne algoritmen består av flere lag av *nevroner* som sender og mottar numeriske verdier til hverandre, basert på vektene de får fra treningsdata. I vårt tilfelle beregner vektene i hvilken grad en kommentar er hatefull eller ikke. Vi har også testet en versjon av BERT-modellen, utviklet av Nasjonalbiblioteket AI Lab (<https://github.com/NBAiLab/notram>). Den modellen gir omtrent en prosent mer i generell nøyaktighet og en bedre «recall» på hatefulle kommentarer. Ulempen er at det gir en betydelig lavere «recall» på «ikke-hatefulle» kommentarer. Ettersom det finnes mange flere ikke-hatefulle kommentarer generelt, overestimerer modellen hatefulle kommentarer i sin klassifisering. I tillegg krever BERT-modellen en *grafikkprosessor* (GPU) som ikke er lett tilgjengelig for folk flest.

¹⁹ Se koding på GitHub for detaljer om tekstprosessering: <https://github.com/upretiramesh/OsloMet-KARTLEGGING-AV-OMFANGET-AV-HATEFULLE-ELLER-DISKRIMINERENDE-YTRINGER>

Vi brukte det annoterte datasettet, som beskrevet over, for å trene modellen til å klassifisere ytringer som hatefulle eller ikke-hatefulle.²⁰ For å trene modellen delte vi det annoterte datasettet i tre sett: et treningssett, et valideringssett og et testsett. Treningssettet består av et tilfeldig utvalg av 70 prosent av det annoterte datasettet, jevnt fordelt mellom hatefulle og ikke-hatefulle ytringer. Treningssettet ble brukt for å lære opp modellen til å gjenkjenne mønstre for hatefulle og ikke-hatefulle ytringer. De to andre datasettene består begge av 15 prosent av de resterende kommentarene. De to små datasettene brukte vi for å validere og teste modellen.²¹

Evalueringen av en slik modell er ofte basert på fire mål: nøyaktighet (*accuracy*), presisjon (*precision*), tilbakekalling (*recall*) og F1-score (Brekke et al., 2019). Hvert av disse målene gir kunnskap om hvordan modellen presterer i klassifiseringen av hatefulle ytringer (*predikert verdi*) sammenlignet med det forskerne kodet selv (*faktisk verdi*). Siden vi bruker en binær klassifikasjon for vår modell, får vi fire mulige resultater som kan representeres i en matrise (*confusion matrix*) (Tabell 1).

Tabell 1

Utfallsrom – relasjon mellom predikerte og faktiske klassifiseringer		
	Hateful full ytring – predikert av modellen	Ikke-hateful full ytring – predikert av modellen
Hateful full ytring – kodet av forskerne	Sanne Positive (SP)	Falske Negative (FN)
Ikke-hateful full ytring – kodet av forskerne	Falske Positive (FP)	Sanne Negative (SN)

Kilde: Tilpasset etter Brekke et al. (2019).

Nøyaktighet måler den generelle prestasjonen av modellen ved å summere alle tilfeller hvor modellen og forskerne har kodet ytringene likt (SP + SN) og deler dette med det totale antallet tilfeller (SP + SN + FP + FN). Begrensningen med dette målet er at det ikke er sensitivt til ubalansert datasett, hvor en kategori har mange flere tilfeller enn den andre. Hvis en modell er god til å klassifisere den kategorien med høyere antall tilfeller, har modellen ofte en høy nøyaktighet selv om den har en dårligere presisjon når det gjelder kategorien med få tilfeller. Siden vi har langt flere tilfeller av ikke-hatefulle kommentarer enn hatefulle kommentarer, må vi se også på andre mål.

De andre to målene bidrar til en mer raffinert evaluering av modellens ytelse innenfor hver kategori (hatefulle eller ikke-hatefulle ytringer). Presisjon angir hvor mange av totalt antall prediksjoner i en kategori som er riktig klassifisert (eller sanne). Det betyr at vi tar antall sanne positive eller sanne negative og deler på summen av deres respektive kolonner. I Tabell 1 er formelen for sanne positive: $SP/(SP+FP)$. Tilbakekalling måler det totale antallet av de sanne tilfellene i en kategori som er riktig predikert, og er en indikator på hvor mange av de sanne tilfellene modellen ikke fanger opp. Vi tar antall sanne positive eller sanne negative og deler på summen av deres respektive linjer, noe som gir oss formelen (for sanne positive): $SP/(SP+FN)$. F1-score kombinerer presisjon og tilbakekalling for å gi en middelværdi av de to målene. F1-score beregnes med følgende formel: $F1 = 2 \times (\text{presisjon} \times \text{tilbakekalling}) / (\text{presisjon} + \text{tilbakekalling})$.

²⁰ Se for eksempel rapport fra Institutt for samfunnsforskning for et overblikk over bruk av maskinlæring og andre metoder i omfangsstudier av hatefulle ytringer (Nadim et al., 2018)

²¹ Prosessering av datagrunnlaget og spesifikasjoner er tilgjengelig i:

<https://github.com/upretiramesh/OsloMet-KARTLEGGING-AV-OMFANGET-AV-HATEFULLE-ELLER-DISKRIMINERENDE-YTRINGER>

Resultatene fra testsettet av den trente modellen er gjengitt i Tabell 2. Modellen vi testet hadde en generell nøyaktighet (*accuracy*) på 81 prosent i sin gjenkjenning av ytringer i testsettet. Når vi ser på hver av kategoriene, hatefulle vs. ikke-hatefulle ytringer, var modellen bedre til å gjenkjenne ikke-hatefulle enn hatefulle ytringer i testsettet. Modellen klarte å fange opp nærmest alle ikke-hatefulle ytringer, siden tilbakekalling var 96 prosent for den kategorien, mens kun 36 prosent av alle hatefulle ytringer ble fanget opp av modellen. Klassifiseringsmodellen gikk altså glipp av en god del hatefulle ytringer i testsettet. Testingen av modellen indikerte dermed at modellen ville gi en underestimering av det totale antallet hatefulle ytringer. Modellens ulikhet med hensyn til tilbakekalling er et resultat av justeringer vi har gjort for å prøve å minimere risiko for overestimering av hatefulle ytringer. Vi ønsket en modell som i størst mulig grad fanger opp hatefulle ytringer og som har færrest mulig «falske positive» for hatefulle ytringer, (dvs. ikke-hatefulle ytringer klassifiserte som hatefulle).

Tabell 2

	Modellens ytelse på testsett			
	Presisjon	Tilbakekalling	F1-score	N
Hatefull ytring	75 %	36 %	48 %	121
Ikke-hatefull ytring	82 %	96 %	89 %	371
Vektet gjennomsnitt	80 %	81 %	79 %	492

Vi testet klassifiseringsmodellen på et randomisert datasett av nye kommentarer som ikke var klassifisert tidligere av forskerne, blant annet for å teste modellens underestimering av hatefulle ytringer. Det randomiserte datasettet besto av en stikkprøve av 200 kommentarer klassifisert av modellen som hatefulle, og 200 kommentarer klassifisert som ikke-hatefulle, basert på vår ordliste. Antall kommentarer var likt fordelt mellom Facebook og Twitter. Det samme datasettet bestående av 400 kommentarer kodet vi manuelt for å sjekke modellens konsistens.

Testingen av modellen på dette randomiserte, uklassifiserte datasettet viste at modellens ytelse generelt var noe dårligere enn for det første testsettet, med henholdsvis 77 mot 81 prosent nøyaktighet (*accuracy*). Resultatene av modellens ytelse på det randomiserte, uklassifiserte datasettet er presentert i Tabell 3. Modellen fanget 79 prosent (tilbakekalling) av alle hatefulle ytringer, og 75 prosent av alle ikke-hatefulle ytringer. Modellens presisjon er imidlertid noe høyere for ikke-hatefulle ytringer enn for hatefulle ytringer. Av alle ikke-hatefulle ytringer identifisert av modellen er 87 prosent av ytringene også kodet av forskerne som ikke-hatefulle, mens tilsvarende andel av hatefulle ytringer er 67 prosent. Testingen av modellen på testsettet og det randomiserte, uklassifiserte datasettet viser altså enkelte forskjeller. Modellens ytelse på uklassifiserte data ser ut til å fange opp flere hatefulle ytringer enn hva testene av testdatasettet antydte, men med den ulempen at en noe høyere andel av ytringene som blir identifisert som hatefulle ytringer, antagelig ikke er hatefulle.

Modellen skal anvendes for å analysere uklassifiserte data fra sosiale medier. Testingen av modellen viser at modellen presterer rimelig bra på nye, uklassifiserte data, og er over verdier vurdert som akseptable for å anvende en slik modell (Vidgen & Yasseri, 2020).

Tabell 3

	Modellens ytelse på uklassifiserte data			
	Presisjon	Tilbakekalling	F1-score	N
Hatefull ytring	67 %	79 %	72 %	169
Ikke-hatefull ytring	87 %	75 %	79 %	231
Vektet gjennomsnitt	78 %	76 %	77 %	400

En kvalitativ analyse av datasettet med uklassifiserte kommentarer viste at modellen har noen kjente utfordringer for automatisert klassifisering av hatefulle ytringer. Å identifisere alle hatefulle ytringer er utfordrende, for eksempel på grunn av problemet med å fange opp ironi, retoriske spørsmål og sitater fra andre som ytrer seg på en hatefull måte (se kunnskapsoppsummeringen). Dette er med på å forklare hvorfor modellen ikke klarer å fange opp alle hatefulle ytringer eller klassifiserer noen ikke-hatefulle kommentarer feil. Denne utfordringen gjør at modellen antagelig kan overestimere omfanget av hatefulle ytringer, siden testingen av modellen indikerer at noen av de identifiserte hatefulle ytringene antagelig ikke er hatefulle (altså falske positive).

Overestimering kan skje med vår modell fordi det finnes mange flere ikke-hatefulle kommentarer. Datasettene våre er altså ubalanserte. Det betyr, for eksempel, at hvis vi har 1000 ikke-hatefulle kommentarer og 100 hatefulle kommentarer, og bruker ytelsen på uklassifiserte data som referanse, så vil vår modell gi oss ca. 21 hatefulle kommentarer klassifiserte som ikke-hatefulle og ca. 250 ikke-hatefulle kommentarer klassifiserte som hatefulle. På den måten får vi en overestimering av hatefulle kommentarer på mer enn 200 prosent. Modellen egner seg derfor ikke til å estimere antall hatefulle kommentarer over en enkeltperiode. Vi bør derfor tolke resultater om absolutt antall hatefulle kommentarer i en enkeltperiode – og særlig perioder med få kommentarer – med forsiktighet. For tendenser over tid har det antagelig mindre betydning. Modellen egner seg godt til å sammenligne utvikling av hatefulle ytringer over tid.

Vi har gode grunner til å tro at vår modell er minst like god til å identifisere relative endringer i antall hatefulle kommentarer over tid som andre modeller. For det første viser vår modell en ytelse som er like god på uklassifiserte data som for eksempel modellen utviklet av Vidgen og Yasseri (2020). For det andre trente vi modellen på data som dekker en mye lengre tidsperiode enn de fleste studiene om hatefulle ytringer. Mens Vidgen og Yasseri (2020) eller Burnap og Williams (2015) begrenser datamaterialet til noen måneder eller noen få år, dekker vi en tiårsperiode (2010–2021). Dette er viktig siden både tematikk og språk relatert til hatefulle kommentarer endrer seg over tid. Ved å trene og teste modellen med data fra en lengre tidsperiode er modellen mindre sårbar for å bli spesialisert på et begrenset fenomen eller en bestemt debatt. I tillegg bruker vi data fra flere kilder. Mens de fleste studiene bruker Twitter-data (Poletto et al., 2017), bruker vi også kommentarer fra Facebook og andre nettsider. Slik har vi en mer variert samling av brukere. I tillegg har vi potensielt lengre kommentarer, siden Facebook ikke har begrensning i antall tegn i kommentarfeltene.

4.1.5 Data

Våre stordata består av innlegg og kommentarer fra Facebook og Twitter for tiårsperioden januar 2010 til september 2021. Facebook og Twitter er populære plattformer i Norge. Det er bakgrunnen for valget av nettopp disse to sosiale mediene. Ifølge IPSOS hadde ca. 3,4 millioner nordmenn en profil på Facebook, mens 1,1 millioner hadde en konto på Twitter i begynnelsen av 2021.²² En

²² <https://www.ipsos.com/nb-no/ipsos-some-tracker-q121>

annen viktig grunn for valget av plattformer er at studien er basert på offentlige innlegg og kommentarer. På Twitter er alle innlegg og kommentarer i utgangspunktet offentlige. På Facebook har vi valgt ut noen offentlige sider som er nærmere beskrevet under. Mange av innleggene og kommentarene på Facebook er ikke offentlige innlegg, men skjer i lukkede private grupper og mellom private profiler. Av hensyn til personvern har vi i denne studien ikke samlet inn private innlegg på Facebook. Det ville dessuten vært betydelige tekniske utfordringer ved å innhente data som ikke er offentlige.

Twitter og Facebook jobber aktivt for å fjerne kommentarer som oppfordrer til vold eller som fremmer hatefulle ytringer på plattformene sine. Nyhetsredaksjoner jobber også aktivt for å kontrollere hatefulle kommentarer på sine Facebook-sider og Twitter-profiler. Det betyr at våre data består av et redigert utvalg kommentarer, og dataene representerer et «redigert» bilde av hatefulle kommentarer de siste ti årene. Carlson og Rousselle (2020) viser for eksempel at av 311 kommentarer på Facebook som ble rapportert som hatefulle, ble rundt halvparten av dem fjernet.

For å kunne identifisere innlegg og kommentarer på Twitter og Facebook som omhandler muslimer, har vi utviklet en ordliste som er direkte eller indirekte relatert til denne gruppen. Søkeordene i ordlisten vår er:

muslimer, muslim, islam, islamisering, snikislamisering, Mohammed, jihad, profet, koran, taqiyya, eurabia, hijab, niqab, pakkis, innvandring, asylsøker, flyktning.

Selv om de fire siste ordene ikke nødvendigvis er relaterte til muslimer, forekommer de ofte i kommentarer som har en forbindelse til hatefulle ytringer mot muslimer. I ordlisten har vi også inkludert forskjellige måter å skrive noen av søkeordene på, som for eksempel Mohammed, Muhammad, Mohamed osv. Basert på denne ordlisten har vi innhentet innlegg og kommentarer på Twitter og Facebook som inneholdt minst ett av disse ordene.

Til denne studien har vi innhentet fire datasett: to fra Twitter og to fra Facebook. Datasettene består av: selve kommentaren, datoen kommentaren ble skrevet og et anonymisert ID-nummer for brukeren som skrev kommentaren.

Disse fire datasettene inneholder samlet sett 800 000 kommentarer om muslimer fra Facebook og Twitter for perioden januar 2010 til september 2021. Det er altså disse 800 000 kommentarene vi har brukt for å identifisere hatefulle og diskriminerende ytringer mot muslimer.

4.1.6 Data fra norske nyhetskanaler (Datasett 1 og 2)

De fleste aviser og nyhetskanaler i Norge bruker Facebook og Twitter for å dele nyhetsoppslagene sine. Sosiale medier har slik blitt et *de facto* kommentarfelt for de fleste nyhetskanaler i Norge, etter at mange stengte ned kommentarfeltene på egne nettsider. De to første datasettene består av kommentarer under innlegg på Facebook-sidene og Twitter-kontoene til noen utvalgte nyhetsmedier. Vi brukte ordlisten vår for å finne relevante saker som omtaler muslimer fra januar 2010 til september 2021. Deretter innhentet vi kommentarene under disse sakene.

Nyhetsmediene vi har hentet kommentarer fra er:

- Aftenposten
- Bergens Tidende
- Dagbladet
- Dagen
- Dagsavisen
- Document.no
- Nettavisen

- NRK
- Resett
- Rights.no
- TV2
- Vårt Land

På Facebook brukte vi CrowdTangle²³ for å finne innlegg fra nyhetsmedier som hadde minst ett ord fra ordlisten vår. Deretter innhentet vi kommentarene under innleggene. For innhenting av kommentarer brukte vi en kombinasjon av manuelle og automatiserte metoder (*web scraping*). Resultatet ble 9297 relevante innlegg fra nyhetsmedier på Facebook, og 296 431 tilhørende kommentarer. På Twitter brukte vi Twitters egen API (Application Programming Interface). Denne gir forskere tilgang til alt innholdet på Twitter. Vi fant 3691 relevante innlegg fra nyhetsmedier på Twitter, og 3382 kommentarer under disse nyhetssakene. Den store forskjellen i antall kommentarer mellom de to plattformene er ikke overraskende. Facebook har flere brukere enn Twitter, og kommentarfeltene under nyhetssakene på Facebook er mer brukt.

4.1.7 Data fra Twitter (Datasett 3)

Vi har også samlet et større datasett fra Twitter som består av alle innlegg som omtaler muslimer fra januar 2010 til september 2021. Dette datasettet har vi også samlet ved å bruke ordlisten vi beskrev tidligere. Det inneholder 296 056 twittermeldinger.

4.1.8 Data fra islamkritiske Facebook-sider (Datasett 4)

Vi var interesserte i å finne ut i hvilken grad det forekommer hatefulle ytringer i såkalte islamkritiske forum. Det fjerde datasettet består derfor av kommentarer på innlegg publisert på Facebook-sidene til tre offentlige profiler/grupper som er kjent for å være kritiske mot Islam. De tre profilene/gruppene er: Stopp Islamiseringen av Norge (SIAN), Norge Fritt for Islam og Hege Storhaug sin offentlige profil. Basert på ordlisten vår innhentet vi først innlegg som omtaler muslimer ved å bruke CrowdTangle. Deretter innhentet vi kommentarer som refererer til disse innleggene. Dette datasettet utgjør 11 361 innlegg og 188 546 tilhørende kommentarer fra oktober 2016 til september 2021.²⁴ Som nevnt tidligere, foregår mange av samtalene og diskusjonene om muslimer på Facebook i lukkede grupper eller på private profiler. De tre sidene er blant de få som offentliggjør slikt innhold på Facebook-sidene sine. Selv om dette er et lite og vilkårlig utvalg, mener vi det kan gi en indikasjon på noe av innholdet i de samtalene som foregår blant dem som beskriver seg selv som islamkritiske. De tre gruppene har til sammen litt i overkant av 88 000 følgere.

4.2 Analyser – Omfang av omtale av muslimer

For å beskrive utviklingen i omfanget av innlegg og kommentarer på Twitter og Facebook som omhandler muslimer, vil vi først gi en analyse av utviklingen i antall nyhetssaker publisert av norske medier. Deretter vil vi beskrive i hvilken grad disse nyhetssakene skaper en reaksjon blant leserne.

4.2.1 Nyhetssaker som kan relateres til muslimer

Vi vil her presentere fordelingen av nyhetssaker som inneholdt minst ett av søkeordene som relateres til muslimer og som ble publisert på Facebook-sidene og Twitter-kontoene til tolv utvalgte

²³ Dette er en tjeneste laget av Facebook for journalister og forskere som har interesse av innholdet på offentlige profiler og i grupper på plattformen. Se mer på www.crowdtangle.com.

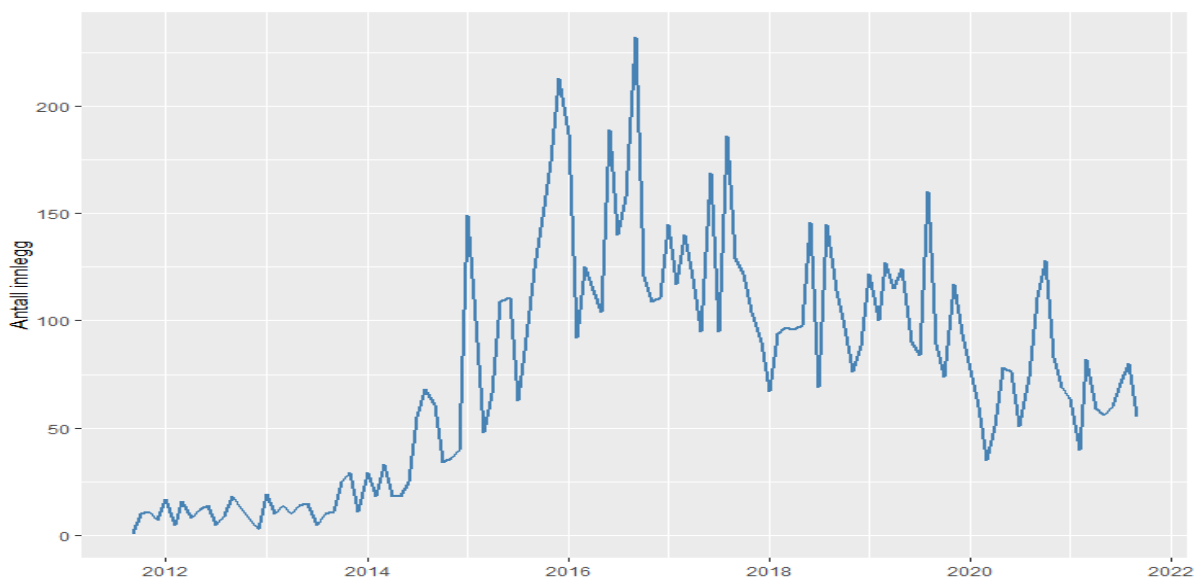
²⁴ Alle de tre sidene begynte sine offentlige aktiviteter mellom oktober og november 2016.

nyhetsmedier i Norge for perioden januar 2010 til september 2021. Her viser vi hvilke nyhetsmedier som omtaler muslimer oftest i løpet av perioden vi analyserer.

Som nevnt i metodekapittelet, omhandler disse nyhetssakene temaer knyttet til muslimer, i tillegg til innvandring, flyktninger og den Islamske Staten (IS). Disse sistnevnte sakene stimulerer ofte til kommentarer om muslimer – og ikke minst hatefulle kommentarer. Totalt har vi samlet 9297 nyhetssaker på Facebook-sidene og 3691 nyhetssaker på Twitter-kontoene til de tolv utvalgte nyhetsmediene i Norge.

Figur 1 viser antall nyhetssaker som omtaler muslimer på alle de tolv utvalgte nyhetsmedienes Facebook-sider per måned for perioden januar 2010 til september 2021. Antall nyhetssaker har hatt en betydelig økning de siste fem årene. I perioden 2010–2015 ble det publisert gjennomsnittlig 41 nyhetssaker per måned, som til sammen gir 2160 nyhetssaker for denne femårsperioden. Mens i perioden 2016–2021 identifiserte vi 103 nyhetssaker per måned, som til sammen gir 7137 nyhetssaker for denne siste femårsperioden. Det representerer en økning på 151 prosent i antall nyhetssaker som omhandler muslimer når vi sammenligner den første og siste femårsperioden. Antallet publiserte nyhetssaker var aller høyest i årene 2015–2017.

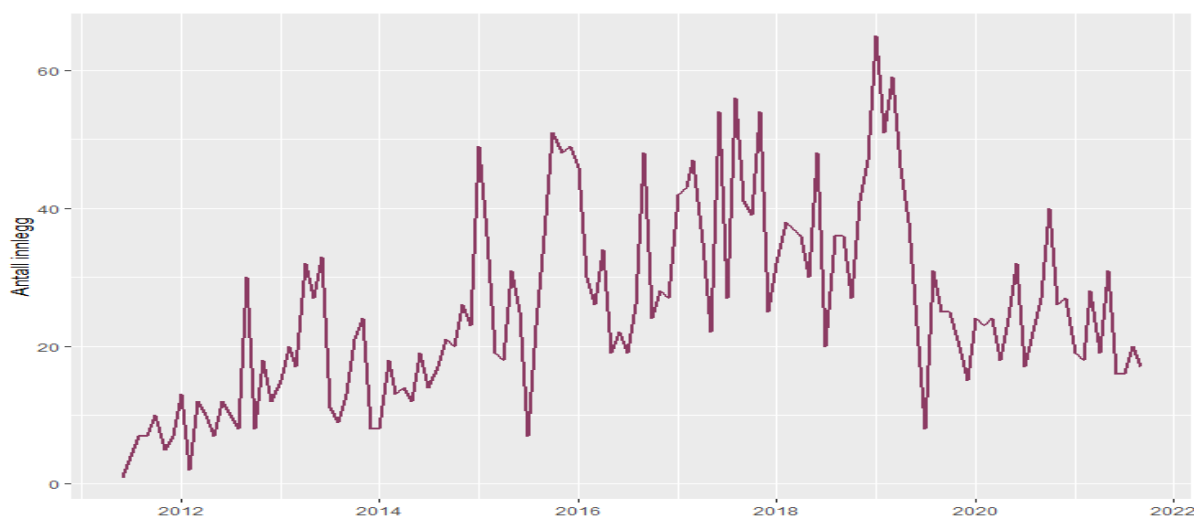
Figur 1 – Totalt antall nyhetssaker på nyhetsmediers Facebook-sider per måned – alle medier



En liknende utvikling har vi funnet på Twitter (se Figur 2), men utviklingen er svakere siden Twitter brukes mindre av nyhetsmediene enn Facebook. På Twitter har de tolv utvalgte nyhetsmediene publisert 18 nyhetssaker per måned i perioden 2010–2015, som til sammen er 1006 nyhetssaker for femårsperioden. For perioden 2016–2021 er antall publiserte nyhetssaker 31 per måned, som til sammen gir 2156 publiserte nyhetssaker for femårsperioden. Det representerer en økning på 78 prosent når vi sammenligner første og siste femårsperiode.

Det er vanskelig å peke på en enkelt årsak som forklarer denne trenden. Denne studien går heller ikke i dybden for å identifisere årsaker. Likevel er flyktningkrisen i Europa og Norge fra 2015 en viktig hendelse som kan forklare økningen (Hognestad & Lamark, 2017). Flyktningkrisen og dens konsekvenser i de følgende årene har skapt en intens offentlig debatt, ikke bare om håndteringen av krisen, men også om innvandring og integrering generelt, hvor muslimer ofte står i sentrum av debatten. En annen forklaring er at den generelle bruken av sosiale medier har økt.

Figur 2 – Antall nyhetssaker på Twitter per måned – alle medier



For perioden 2010–2021 har vi sett en markant endring i hvilke nyhetsmedier som publiserer nyhetssaker som omhandler muslimer. For å presentere endringene for periodene 2010–2015 til 2016–2021 har vi innenfor hver av periodene gruppert de mediene som har høyest andel publiserte nyhetssaker om muslimer og som til sammen utgjør mer enn 50 prosent.²⁵

I løpet av perioden 2010–2015 sto tradisjonelle medier som Aftenposten, Dagen, Dagbladet og Vårt Land for 56 prosent av alle nyhetssaker som omtalte muslimer, mens nyere medier som Resett, Rights.no, Document.no og Nettavisen publiserte 54 prosent av alle nyhetssaker på Facebook som omtalte muslimer i perioden 2016–2021. Det samme ser vi på Twitter. I løpet av 2010–2015 publiserte Dagen, Document.no og Aftenposten 57 prosent av alle nyhetssaker om muslimer på Twitter, mens i perioden 2016–2021 publiserte Resett, Document.no og Rights.no 64 prosent av alle nyhetssakene som omtaler muslimer.

4.2.2 Reaksjoner på nyhetssakene

Like viktig som å anslå antall saker som kan relateres til muslimer, er å se på i hvilken grad sakene skaper en reaksjon blant leserne. Selv om det har vært en økning i antall saker, kunne man tenke at økningen ikke nødvendigvis skaper mer oppmerksomhet hos brukerne av plattformene.

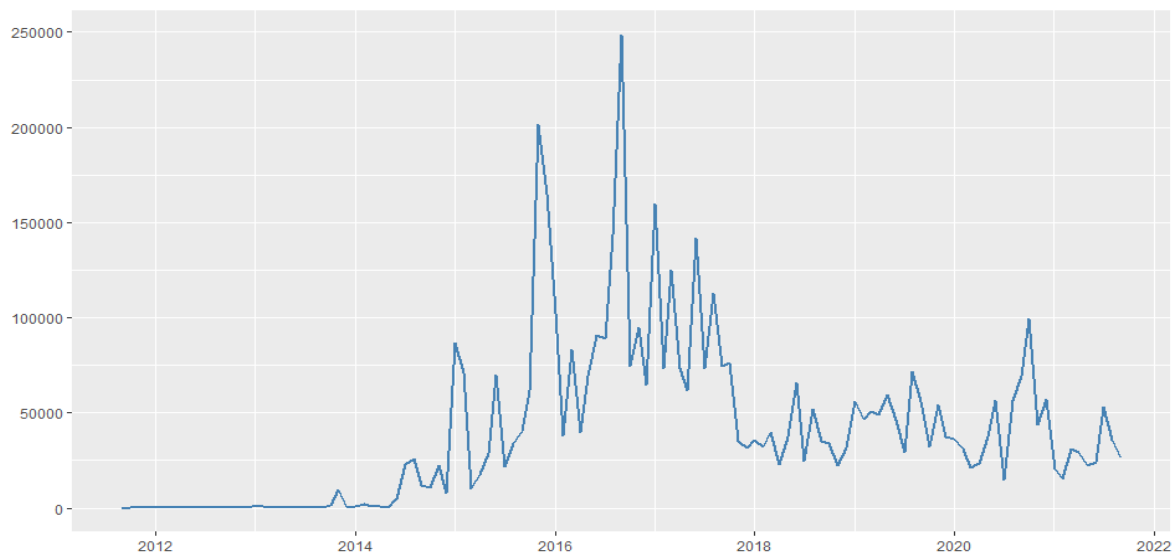
På Facebook måler vi brukeraktiviteten ved å legge sammen alle «likes», delinger og kommentarer.²⁶ På Twitter måler vi brukeraktiviteten ved å legge sammen alle kommentarer, «retweets» og «likes». Det generelle bildet er at vi ser en økning av interaksjoner på både Facebook og Twitter i perioden 2016–2021, sammenlignet med perioden 2010–2015. Figur 3 viser utvikling av brukeraktivitet på Facebook i løpet av tiårsperioden vi analyserer.

I perioden 2010–2015 var det registrert 18 095 interaksjoner per måned på nyhetssaker som omhandler muslimer, mens i 2016–2021 var det registrert 58 000 interaksjoner per måned. Sammenligner man de to periodene, er økning i antall interaksjoner på nyhetssaker som omhandler muslimer på hele 220 prosent. En stor del av økningen er antagelig et resultat av økende popularitet av plattformen i Norge.

²⁵ Disse kriteriene for gruppering vil bli anvendt i de påfølgende analysene av endringer mellom nyhetsmedier.

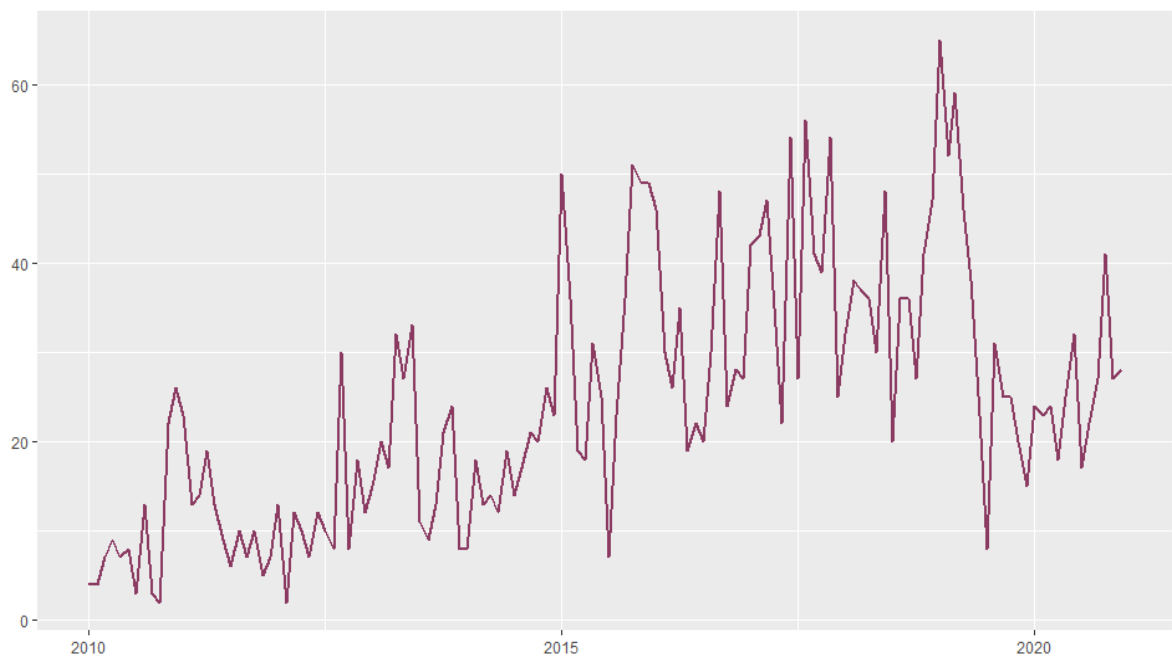
²⁶ Vi har også inkludert alle nye «reaksjonsknapper» som Facebook gjorde tilgjengelig i 2016 og 2020.

Figur 3 – Totalt antall brukeraktiviteter på Facebook per måned – alle medier



Figur 4 viser utvikling av brukeraktivitet på Twitter i samme periode. Vi finner en økning på 773 prosent mellom de to periodene, men her var startpunktet mye lavere, med 26 interaksjoner per måned i 2010–2015 mot 227 interaksjoner per måned i 2016–2021. Den betydelige forskjellen i omfanget interaksjoner på Facebook og Twitter viser hvordan Twitter er både mindre populær enn Facebook og mindre relevant for nyhetskanaler

Figur 4 - Antall brukeraktivitet på Twitter per måned – alle medier



Vi finner også en endring med hensyn til hvilke mediekanaler som får flest interaksjoner på nyhetssaker om muslimer. I perioden 2010–2015 var 60 prosent av alle interaksjoner på nyhetssaker om muslimer på Facebook-sidene til TV2, Dagbladet og Nettavisen. I perioden 2016–

2021 var 60 prosent av interaksjonene knyttet til nyhetssaker publisert av nyere medier som Resett, Rights.no, Nettavisen og Document.no. På Twitter fant vi en lignende trend. I perioden 2010–2015 utgjorde interaksjoner på nyhetssaker publisert bare av Aftenposten og VG 56 prosent av alle interaksjoner på nyhetssaker publisert av de tolv utvalgte mediene. I perioden 2016–2021 utgjorde nyhetssaker publisert bare av Rights.no og Resett 63 prosent av alle interaksjoner på nyhetssaker.

Endringene i hvilke mediekanaler som er de mest aktive i sosiale medier når det gjelder å omtale muslimer og islam, er en viktig bakgrunn for å forstå trendene vi ser over tid om hatefulle og diskriminerende ytringer i forbindelse med nyhetssaker.

4.2.3 Hatefulle kommentarer i forbindelse med nyhetssaker

For å vurdere omfanget av hatefulle og diskriminerende ytringer mot muslimer i forbindelse med nyhetssaker, analyserte vi alle kommentarene vi hadde samlet i forbindelse med nyhetssaker om muslimer i vår klassifiseringsmodell. Av de totalt 296 626 kommentarene på Facebook, publisert på Facebook-sidene til de tolv utvalgte nyhetsmediene i perioden januar 2010 til september 2021, ble 3,7 prosent klassifisert som hatefulle eller diskriminerende ytringer. På Twitter var andelen noe høyere, da 6,5 prosent av alle kommentarene på nyhetssaker ble klassifisert som hatefulle eller diskriminerende ytringer. Det totale antallet kommentarer på Twitter er imidlertid betydelig lavere enn på Facebook. For tiårsperioden januar 2010 til september 2021 var antall twittermeldinger 1770. Både på Facebook og Twitter er omfanget av hatefulle kommentarer generelt sett lavt sett i forhold til alle kommentarer som skrives.

Vår modell kan overestimere antallet hatefulle kommentarer i enkelte perioder, som allerede nevnt i metodekapittelet. Dette har imidlertid mindre å si for den tydelige økningen i omfanget av hatefulle kommentarer de siste fem årene, både på Facebook og på Twitter. På Facebook fant vi 35 hatefulle kommentarer i forbindelse med publiserte nyhetssaker per måned i perioden 2010–2015, og 132 hatefulle kommentarer per måned i perioden 2016–2021. På Twitter ble bare 116 kommentarer klassifiserte som hatefulle for hele perioden 2010–2021. Men 113 var i perioden 2016–2021, mot kun 3 hatefulle kommentarer for perioden 2010–2015.

Analysene våre viser også at de fleste hatefulle kommentarene er skrevet i forbindelse med nyhetssaker publisert av noen få medier. På Facebook var 53 prosent av alle hatefulle eller diskriminerende kommentarer skrevet i forbindelse med nyhetssaker publisert av Nettavisen og TV2 i perioden 2010–2015. Dette endret seg for perioden 2016–2021, da 51 prosent av alle hatefulle kommentarer ble skrevet i forbindelse med saker fra Resett og Rights.no. På Twitter ble 68 prosent av alle 116 hatefulle kommentarer i perioden 2010–2021 skrevet i forbindelse med saker publisert av Rights.no og Resett.

Det kan være flere grunner til korrelasjonen mellom nyhetssaker fra disse alternative mediene og hatefulle og diskriminerende kommentarer. En forklaring kan være den veldig kritiske vinklingen mot muslimer som disse mediene fremmer, og at vinklingen i seg selv kan inspirere til hatefulle og diskriminerende kommentarer. En annen forklaring kan ha sammenheng med i hvilken grad disse mediene kontrollerer og redigerer kommentarfeltene. En tredje forklaring kan handle om de ulike medienes lesere og deres holdninger til islam og muslimer. Vi går ikke nærmere inn på disse forklaringsmulighetene i denne rapporten.

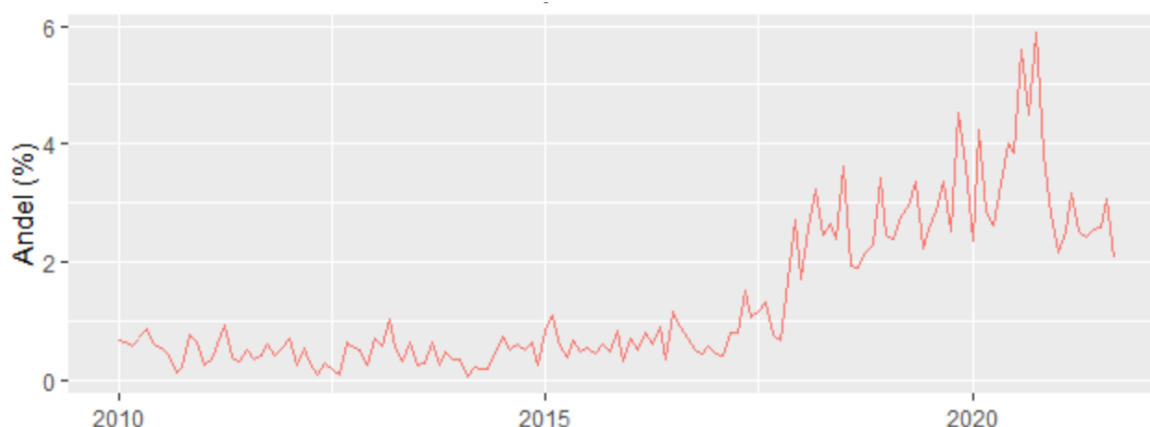
4.2.4 Hatytringer mot muslimer i debatter og samtaler på Facebook og Twitter

Nyheter og kommentarer knyttet til nyhetssaker utgjør bare en veldig liten del av all den aktiviteten som foregår på sosiale medier. Vi har derfor også samlet kommentarer som ikke er direkte relaterte til nyhetssaker, for å vise noe av den større samtalen som foregår i sosiale medier. Som beskrevet

i metodekapittelet har vi brukt den samme ordlisten for å identifisere relevante nyhetssaker for å innhente twittermeldinger som relateres til muslimer, mens på Facebook har vi samlet kommentarer på innlegg fra et utvalg av offentlige grupper/profiler.

I disse analysene finner vi den samme økende trenden som i kommentarfelter under nyhetssaker. Basert på de 296 056 kommentarene innhentet fra Twitter for hele tiårsperioden, viser Figur 5 en tydelig økning de siste årene. Andelen hatefulle kommentarer av alle twittermeldinger per måned har økt fra 0,4 prosent i perioden 2010–2015 til 2,2 prosent i perioden 2016–2021. Andelen hatefulle kommentarer er imidlertid rundt seks prosent i enkelte perioder, som september–november 2020.

Figur 5 – Andel i prosent av hatefulle kommentarer på Twitter



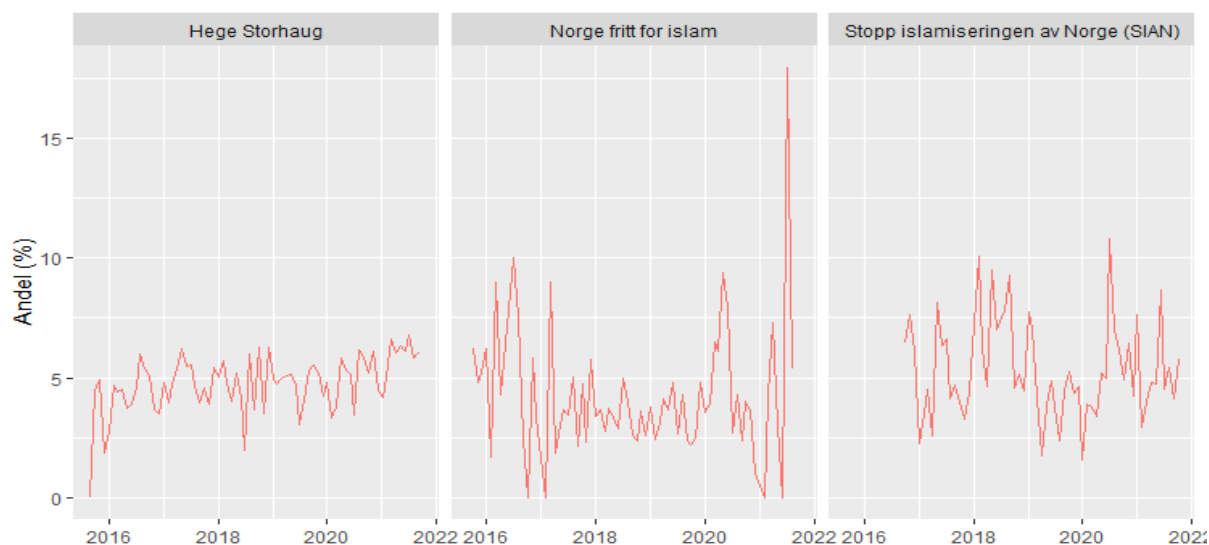
Vi leste kommentarer i perioden med høy andel hatefulle kommentarer, men klarte ikke å identifisere et hovedtema. Det kan spekuleres i at drapet på den franske læreren som viste Muhammed-tegninger til klassen utløste denne trenden, men for å kunne gi noen sikrere svar er det behov for en mer detaljert studie av hendelser som kan påvirke debattene i sosiale medier.

De tre Facebook-sidene SIAN, Norge Fritt for Islam og Hege Storhaug er kjente for å være svært kritiske mot Islam. Det eksisterer imidlertid en betydelig forskjell i antall kommentarer som er publisert på disse Facebook-sidene. Av alle 188 359 kommentarer innhentet fra disse gruppene, kommer 65 prosent fra Hege Storhaugs Facebook-side, 20 prosent fra Norge Fritt for Islam og 15 prosent fra SIAN. Fordelingen reflekterer til en viss grad antall følgere som sidene har. Hege Storhaugs Facebook-side er mest populær, med ca. 45 000 følgere, mens Norge Fritt for Islam har ca. 31 000 og SIAN har ca. 11 500 følgere.

På disse Facebook-sidene ser vi en betydelig høyere andel av hatefulle og diskriminerende kommentarer enn på Facebook-sidene til nyhetsmediene og på Twitter. Figur 6 viser at andel hatefulle kommentarer generelt ligger høyere enn på Twitter. Av kommentarer innhentet fra disse gruppene siden de ble opprettet, ble 9165 kommentarer, 4,8 prosent, klassifisert som hatefulle og diskriminerende. Andelen av hatefulle og diskriminerende kommentarer per måned varierer imidlertid kraftig. Enkelte måneder overstiger andelen 15 prosent, som figur 6 viser.²⁷

²⁷ Ettersom alle de tre sidene ble aktive enten i oktober eller november 2016, har vi ikke en lang nok tidsserie til å kunne vite om det var en økning i andelen av hatefulle kommentarer.

Figur 6 – Andel i prosent av hatefulle kommentarer på utvalg av Facebook-sider



Det er også interessant å undersøke hvem som skriver disse hatefulle kommentarene. Våre analyser viser at de aller fleste ikke ytrer seg veldig ofte på en hatefulle måte, og at en stor andel av hatefulle kommentarer er konsentrert i en liten gruppe av profiler som er ekstremt aktive. På Facebook, for eksempel, identifiserte vi 3855 profiler som har skrevet minst én hatefulle kommentar på Facebook-sidene til de tre gruppene. De 10 prosent mest aktive profilene, som representerer 378 profiler, sto bak nesten 45 prosent av alle de hatefulle kommentarene vi innhentet. Når vi snevrer analysen inn til å gjelde 1 prosent av de mest aktive profilene, står de bak 16 prosent av alle de hatefulle kommentarene.

På Twitter er denne tendensen enda tydeligere. Av alle hatefulle kommentarer innhentet med ordlisten, er de 1 prosent mest aktive profilene (n=14) ansvarlige for nesten 28 prosent av de hatefulle ytringene. Det betyr at en liten gruppe står bak mye av aktiviteten, og at disse i stor grad er ansvarlige for økningen i hatefulle kommentarer vi fant på Twitter. Mens de 10 mest aktive profilene i perioden 2010–2015 sto bak 133 hatefulle kommentarer, sto de 10 mest aktive profilene bak 1028 hatefulle kommentarer i perioden 2016–2021. Disse veldig aktive profilene er ikke de samme i de to periodene. Vi har også brukt tre forskjellige «bot detectors» for å sjekke om profilene er drevet av mennesker og ikke roboter. Alle profilene besto testene og virker å være reelle kontoer som virkelige personer står bak.²⁸

4.3 Oppsummering

I denne studien har vi benyttet stordata for å si noe om omfanget av hatefulle og diskriminerende ytringer mot muslimer over tid. Vi har samlet mer enn en million offentlige kommentarer som omhandler muslimer på Facebook og Twitter. I tillegg har vi utviklet og trent en klassifiseringsmodell, basert på en kunstig nevralt nettverksalgoritme, til å identifisere hatefulle ytringer og å skille hatefulle ytringer fra ikke-hatefulle ytringer på norske sosiale medier. Ved hjelp av våre stordata har vi sammenlignet antallet nyhetssaker som omhandler muslimer i periodene

²⁸ Vi brukte disse tjenestene: 1) Bot Sentinel (www.botsentinel.com), 2) Botometer (www.botometer.osome.iu.edu), og 3) tweetbotornot (<https://tweetbotornot.mikewk.com>).

2016–2021 og 2010–2015 på norske nyhetsmedier sine Facebook- og Twitter-sider. I analysene våre finner vi en generell økning av saker som omhandler muslimer på disse Facebook- og Twitter-sidene. Vi finner også at det i stor grad er nye alternative medier som står bak denne økningen. Det er også på nye alternative medier vi finner flest hatefulle og diskriminerende kommentarer under nyhetssaker som omhandler muslimer.

Våre funn viser også at andelen hatefulle og diskriminerende kommentarer mot muslimer er lav sett i forhold til alle kommentarer som offentliggjøres daglig på Facebook og Twitter. Vi ser imidlertid en tydelig økning av hatefulle og diskriminerende kommentarer mot muslimer over tid. Dette er en økning vi ser både blant kommentarer relatert til nyhetssaker og i vårt utvalg av kommentarer skrevet i andre debatter og samtaler på Twitter og Facebook. Når vi ser nærmere på profilene som publiserer disse hatefulle og diskriminerende kommentarene, viser det seg at det er et veldig lite antall brukere som står bak de aller fleste av dem.

Modellen vi har utviklet er den første norskspråklige modellen, og den kan også være til nytte for andre som vil analysere norske hatefulle kommentarer. Denne norskspråklige modellen kan også videreutvikles og forbedres eller brukes som en referanse for å teste andre alternative modeller.

5 Analyser av innholdet i hatefulle og diskriminerende ytringer

Vi har gjort noen kvantitative og noen kvalitative analyser av innholdet i hatefulle og diskriminerende ytringer mot muslimer. Disse er nærmere beskrevet under egne overskrifter.

5.1 Kvantitative analyser

De kvantitative analysene er basert på en tema-modellerings algoritme («topic-modelling»), som er en type uovervåket metode for maskinlæring («unsupervised machine learning») hvor en algoritme identifiserer forhold mellom ord og uttrykk i et korpus (i.e. samling av tekster/dokumenter). For den analysen brukte vi Latent Dirichlet Allocation (LDA), en veletablert algoritme for modellering av temaer (Grimmer & Stewart, 2013; Enarsson & Lindgren, 2019). Korpuset vi har analysert ved hjelp av denne tema-modelleringen er de ytringene vi klassifiserte som hatefulle i datasett 3 og 4. Hensikten var å få nærmere innsikt i innholdet til disse hatefulle og diskriminerende ytringene i debatter og samtaler på de to plattformene. Analysene er gjort på Facebook og Twitter.

Som et første steg i analysen, har vi laget ordskyer av de 40 mest frekvente ordene i hvert datasett. Ordene skrevet med størst skrift er de som forekommer hyppigst. Ordskyen til venstre under er fra Twitter, og ordskyen til høyre er fra Facebook.

Figur 7 - Ordsky fra Twitter 30.11.21 (Datasett 3)

Figur 8 - Ordsky fra Facebook 30.11.21 (Datasett 4)



I begge skyene kan vi se at ordene ikke, islam, muslimer og Norge forekommer. Denne kombinasjonen kan vi fortolke som en generell negativitet mot islam og muslimer og Norge. Selv om ordskyene er ganske like, ser vi noen forskjeller. I skyen til venstre, fra Twitter, ser vi at det finnes et større utvalg av temaer rundt de mest fremtredende ordene som islam, muslimer og Norge. Hijab, sharia, barn, kvinner og terror er eksempler på ord som er mer fremtredende i de hatefulle kommentarene fra Twitter. Selv om vi ser noe lignende i ordskyen til høyre, fra Facebook, er ord som politikere, Norge og landet mer fremtredende her enn på Twitter.

Forskjellene i ordskyene ble tydeligere da vi brukte algoritmen for å finne underliggende sammenhenger mellom ordene i kommentarene, altså i arbeidet med å analysere frem temaer. I analysene våre valgte vi å sortere ordene under fem temaer etter å ha testet litt forskjellige antall temaer. Det var de fem temaene som ga oss den mest meningsfylte klassifiseringen.²⁹ I denne analysen representerer hvert tema en samling av forskjellige ord som har høyest sannsynlighet for å forekomme sammen i kommentarene. I tabellene nedenfor viser vi, for hvert tema, den indre sammenhengen (*coherence*)³⁰, utbredelse (*prevalence*)³¹ og de ti mest vanlige felles ordene. For validering og tolkning leste vi stikkprøver av kommentarer klassifisert under hvert tema.

For Twitter, blir det veldig tydelig i tabell 4 at det temaet som har den høyeste indre sammenhengen (Tema 2), er relatert til debatten om bruk av hijab. Selv om det temaet er det minst utbredte i datasettet, tilhører rundt tolv prosent av de hatefulle kommentarene denne debatten. Tema 4, derimot, er det mest utbredte temaet, men er også det temaet med lavest indre sammenheng. Det representerer en rekke heterogene kommentarer som refererer til muslimer, sharia og islam som noe generelt negativt for Norge og nordmenn.

²⁹ I denne metoden finnes det ikke et riktig antall forhåndsbestemte temaer. Forskerne må validere temaene algoritmen identifiserer som meningsfull eller ikke, i etterkant.

³⁰ Indre sammenheng (*coherence*) indikerer sannsynlighet for at noen ord i et tema opptrer sammen i et dokument eller i en tekst.

³¹ Utbredelse (*prevalence*) er andel dokumenter som er klassifisert under et tema.

Tabell 4 – Tema-modellering for Twitter

temaer	indre sammenheng	utbredelse (%)	hovedbegreper
1	0.014	23.27	islam, ikke, land, europa, vestlige, mer, norge, muslimer, må, verdier
2	0.021	12.29	islam, ikke, hijab, gjøre, kvinner, muslimske, norske, barn, ingenting, norge
3	0.016	13.21	islam, ikke, muhammed, ideologi, koranen, drepe, muslimer, hører, vantro, hjemme
4	-0.033	35.14	muslimer, ikke, norge, islam, norske, må, går, sharia, landet, nordmenn
5	-0.011	16.1	islam, norge, europa, ikke, må, politikere, alt, står, drap, våre

Når vi ser på disse temaene fra Twitter, og hvilke ord som forekommer i disse, så ser vi at temaene ser ut til å være relaterte til negativt ladete samtaler om islam, knyttet til:

- Europa, vestlige, Norge, muslimer, verdier (Tema 1)
- hijab, kvinner, muslimske, norske, barn, Norge (Tema 2)
- profeten Muhammed, ideologi, Koranen, dra, muslimer, vantro, hjemme (Tema 3)
- muslimer, Norge, norske, sharia, landet, nordmenn (Tema 4)
- Norge, Europa, politikere, drap, våre (Tema 5)

Folk som skriver hatefulle ytringer på Twitter, ser med andre ord ut til å være opptatt av:

- Europa, det vestlige og verdier
- hijab, kvinner og barn
- islam som religion
- muslimer i Norge og sharia
- politikere i Norge og Europa

Tema-modellering for Facebook viser at bare 5 prosent av hatefulle kommentarer faller under Tema 2, mens Tema 3 og 4 dekker nesten 60 prosent av kommentarene (Tabell 5).

Tabell 5 – Tema-modellering for Facebook

temaer	indre sammenheng	utbredelse (%)	hovedbegreper
1	0.032	19	ikke, norge, land, må, norske, landet, hører, kommer, lover, kultur
2	0.051	5	muslimer, europa, land, ikke, islam, muslimske, muslimene, verden, vantro, muslimske_land
3	0.026	29	ikke, politikere, norge, må, landet, norske, våre, vårt, land, europa
4	0.03	31	ikke, jo, mer, norge, alt, ser, får, går, kommer, må
5	0.042	16	islam, ikke, koranen, norge, muslimer, må, ideologi, europa, religion, islamske

Når vi ser på disse temaene fra Facebook, og hvilke ord som forekommer i disse, så ser vi at temaene ser ut til å være relaterte til negativt ladete samtaler om:

- Norge, det norske, lover og kultur (Tema 1)
- islam, muslimer i Europa og muslimske land (Tema 2)
- politikere, Norge, det norske, vårt land og Europa (Tema 3)
- Norge (Tema 4)
- islam, Koranen, Norge, muslimer, ideologi, Europa, religion og det som er islamsk (Tema 5)

Folk som skriver hatefulle ytringer på Facebook, ser med andre ord ut til å være opptatt av:

- norske lover og kultur
- Europa og muslimske land
- Politikere og deres handlinger
- Norge
- Religionen islam og det islamske

Interessant nok faller disse funnene i stor grad sammen med funn fra en undersøkelse av høyreekstremisme på Twitter i sju europeiske land, som også viser samtaler preget av en generell negativitet (Nilsen et al., 2020).

5.2 Kvalitative analyser

Gjennom kommunikasjon på nettet har hatefulle ytringer tatt en multimodal vending der ulike uttrykksmåter, som for eksempel tekst, bilder og lyd, virker sammen for å skape et meningsinnhold. Denne multimodale vendingen gjør hatefulle ytringer mer effektive, i den forstand at de kan virke mer overbevisende og troverdige. Hatefulle ytringer er dessuten, som all annen språkbruk, stadig i utvikling med nye symboler, memmer,³² uttrykk, metaforer, forkortelser, akronymer og konspirasjonsteorier.

Som nevnt i innledningen inkluderer ikke hatefulle og diskriminerende ytringer på nettet bare skriftlige tekster, men også bilder, symboler og andre ikke-tekstlige ressurser, såkalte sammensatte eller multimodale ytringer (jfr. Nilsen, 2012; Kress, 2010). Vi har derfor også gjort noen kvalitative tekstanalyser for å gi noen inntrykk av hvordan disse multimodale (kombinasjoner av skrift, bilder og symboler) hatefulle og diskriminerende ytringene mot muslimer kan manifestere seg på nettet. Disse kvalitative tekstanalysene er gjort med utgangspunkt i multimodale perspektiver (Kress, 2010). Multimodalitet er en tverrfaglig tilnærming som vektlegger kommunikasjon som mer enn bare språk og skriftlig og verbal interaksjon. En viktig teoretisk forutsetning i multimodalitet er at kommunikasjon trekker på flere modaliteter, for eksempel de skriftlige, visuelle, verbale og ikke-verbale modalitetene, som alle bidrar til meningsskaping. Den multimodale tilnærmingen vektlegger samarbeid mellom ulike modaliteter, for eksempel i plakater, der vi gjerne ser tekst og fotografier som sammen skaper et meningsinnhold i plakaten (Nilsen 2014a).

Analysene av innlegg fra SIAN og Norge fritt for islam sine Facebook-sider viser at disse ikke bare er islamkritiske, slik de selv hevder, men at plakaten og tekstene vi har analysert også er godt egnet til å rakke ned på muslimer og spre negative holdninger mot, følelser om og oppfatninger av muslimer. Dette materialet faller således inn under vår definisjon av hatefulle og diskriminerende ytringer. I det som følger viser vi noen eksempler på plakater og tekster om muslimer som har vært delt på Facebook-sida til Norge fritt for islam og SIAN og som fortsatt ligger der i november 2021. Disse plakaten er gjerne såkalte sammensatte tekster, med bilder og tekst som sammen presenterer ett eller flere budskap, slik som plakaten under.

³² Memer er enkeltideer som blir spredd og utvikler seg i en befolkning gjennom gjentakelser og herming. <https://no.wikipedia.org/wiki/Mem>

Figur 9 – Plakater fra Norge fritt for islam sin nettside



De to plakatenes over er hentet fra Norge fritt for islam sin nettside. Det mest fremtredende virkemiddelet i plakaten til venstre er den hvite teksten: Full stopp av muslimsk innvandring til Norge? I denne setningen ligger det en forhåndsantagelse, nemlig at det pågår en muslimsk innvandring til Norge. Under ser vi en rød tekst: Hvor mange vil svare ja? På fotografiet i bakgrunnen ser vi to nikabkledde kvinner. Den ene kvinnen har også en nikab med et symbol i pannen som mange mennesker vil forbinde med den islamske staten og terrorisme. Når vi sammenstiller teksten med fotografiet, kan plakaten fortolkes som en oppfordring til å svare nei til full muslimsk innvandring til Norge, for muslimsk innvandring er likestilt med nikabkledde kvinner og terrorisme. Disse negative stereotypiene brukes altså til å spre et usant bilde av muslimer, for de fleste muslimske kvinner i Norge kler seg ikke i nikab, og de aller færreste – et veldig lite mindretall – støtter den islamske staten eller andre terrororganisasjoner. Plakaten er godt egnet til å skape et fiendebilde ved å vekke frykt og mistenksomhet overfor muslimerne. Folk som leser plakaten kan få et inntrykk eller et forsterket inntrykk av muslimer som farlige, at alle muslimer er å anse som ekstremister og terrorister. På den måten kan plakaten ramme muslimeres status og anseelse blant leserne av plakaten. Plakaten kan også leses som et argument mot muslimsk innvandring: Vi ønsker ikke muslimsk innvandring, fordi muslimsk innvandring er ensbetydende med nikabkledde kvinner og terrorisme. Plakaten kommuniserer altså noe om hvordan muslimer er, og dette bildet av muslimske kvinner med nikab kan bidra til å forsterke eller opprettholde oppfatninger av muslimer som ekstremister og terrorister. Denne implisitte sammenhengen mellom fotografi og tekst må leseren selv slutte seg til gjennom sin kunnskap om sammensatte tekster. Leseren inviteres til å generalisere over muslimerne gjennom teksten og fotografiet, og ved å motta denne invitasjonen står leseren i fare for å la seg påvirke av hatpratet gjennom verbale og visuelle virkemidler (Nilsen, 2014b).

Plakaten under er hentet fra SIAN sin nettside, og den formidler noe av det samme innholdet som i Norge fritt for islam sine plakater.

Figur 10 – Plakat fra SIAN sin nettside



Også på denne plakaten er muslimske kvinner fremstilt som nikabkledde. Her er ikke nikaben markert med IS-symbolet, men over bildet er det skrevet i rødt, en farge vi gjerne forbinder med advarsler, «vi er moderate». Plakaten kan fortolkes som et argument mot påstander om at det finnes moderate muslimer, for muslimske kvinner kler seg i nikab. Den kan også leses som en advarsel mot å tro at muslimer kan være moderate, og folk som ikke er moderate er gjerne ekstreme, eller til og med voldelige ekstremister.

På SIAN sin Facebook-side finner vi også andre plakater som er godt egnet til å spre mistenksomhet og frykt overfor muslimer, slik vi viser i eksempelet under, som er en mer kompleks sammensatt tekst. Plakaten er delt i to. Øverst ser vi SIAN sin logo og en oppfordring til leserne om å melde seg inn i organisasjonen. Under kan vi lese: Bli med og kjemp! I denne oppfordringen om å bli med og kjempe er det underforstått at det pågår en viktig kamp – en kamp mot islamiseringen av Norge. I den nederste delen får vi svar på hvem kampen drives mot: det islamske brorskapet osv. som jobber bak lukkede dører, altså i skjul, i moskeer støttet av staten. I denne nederste delen er den gule teksten formulert som et retorisk spørsmål: Hva tror dere IRN det islamske brorskapet osv. holder på med bak lukkede dører i sine statsstøttede moskeer! Driver strikkeklubber?

Figur 11 – Plakat fra SIAN sin Facebook-side



Det retoriske spørsmålet er imidlertid ikke egentlig et spørsmål, selv om det har spørsmålets grammatiske form. Retoriske spørsmål er indirekte påstander og kan være et kraftfullt virkemiddel.

De uttrykker en implisitt påstand (Schmidt-Radefeldt, 1977). Et retorisk spørsmål krever ingen svar (Hudson, 1975, s. 16). Svaret er innbakt i ytringen. Det retoriske spørsmålets oppgave er ikke å søke informasjon, slik oppgaven til andre typer spørsmål gjerne er. Det retoriske spørsmålets oppgave er å fremme en påstand og få tilhørerne til å slutte seg til det selvfølgelig i påstanden (Ilie, 1999, s. 128) eller det tilsynelatende selvfølgelig. Det retoriske spørsmålets funksjon er å skape et felles utgangspunkt. Dette utgangspunktet finner man ved å gjøre to ting:

1. Det første man må gjøre, er å omformulere spørsmålet til en påstand. Da får man i dette tilfellet følgende påstand: *(Det islamske brorskap osv.) driver strikkeklubber.*
2. Det neste steget er å negere påstanden, altså å gjøre den negativ ved å sette inn ikke. Da får vi det *egentlige* innholdet i det retoriske spørsmålet: *(Det islamske brorskap osv.) driver ikke strikkeklubber.*

Plakaten er godt egnet til å skape mistenksomhet overfor muslimer i moskeene, da den kan fortolkes som at det drives lyssky virksomhet i moskeene, ikke noe uskyldig som strikkeklubber.

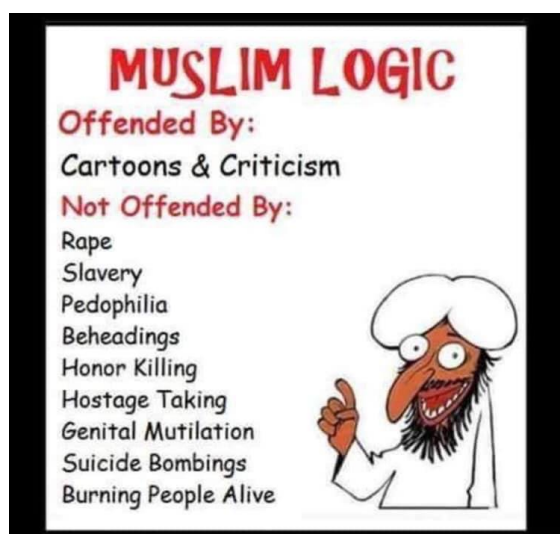
På SIAN sin Facebook-side finner vi også denne plakaten, som vist under. Plakaten kan fortolkes som en beskrivelse av forskjellen på muslimske og andre kvinner. Den kan leses som en visualisering av oss versus dem, der virkemiddelet kontrast brukes som et bevis for hvor forskjellige muslimer er fra andre. Vi ser kontrasten i klesdraktene, og den forsterkes ytterligere i den hvite teksten som er skrevet på svart bakgrunn under bildet. På den måten kan plakaten gi assosiasjoner til ideer om at muslimer og nordmenn ikke kan leve sammen – at de to folkegruppene er for forskjellige. Den kan også leses som at alle muslimske kvinner bærer nikab og er strengt kledd i svart med svarte hansker og svarte sko. I bildet ser vi også at kvinnene går hver sin vei, noe som er med på å forsterke teksten under fotografiet.

Figur 12 - Plakat fra SIAN sin Facebook-side



I hatpratet mot muslimer finner vi også påstander om at muslimer dyrker vold, som vist i kunnskapsoppsummeringen og i de kvantitative analysene, og slik som i plakaten under, som er lagt ut på Facebook-sida til SIAN.

Figur 13 – Plakat fra SIAN sin Facebook-side



På Facebook-sida til SIAN ligger det også noen tekstfiler. En av disse tekstene har tittelen M U H A M M E D – Muslimenes onde forbilde. I teksten argumenteres det slik: «For å få et objektivt sett korrekt bilde av islam må vi se på han som fant opp islam, og utnevnte seg selv til dens siste profet. 91 koranvers sier at Muhammed er det beste eksempel og rollemodell for muslimene. Muhammed var et monster. Derfor er alle gode muslimer dårlige medmennesker.» I denne teksten kan vi også lese: «Idealmuslimen Muhammed likte best kvinner som ikke protesterte når han sodomerte dem.» Leserne får altså vite at slik er den såkalte idealmuslimen. Dette er muslimenes forbilde.

I en annen tekst omtales muslimer som rapefugees, slik du ser i teksten under fra SIAN sin Facebook-side. Teksten og tegningen av mennene med våpen som løper etter en kvinne kan fortolkes som at muslimene ikke er flyktninger, refugees. I henhold til en slik fortolkning er muslimene ikke kommet til Norge fordi de er forfulgt. De er kommet for å voldta norske kvinner.

Figur 14 – Tekstfil fra SIAN sin Facebook-side

ISLAM
Seksualpredatorenes
himmelrike på jord



Islams kvinnefiendtlighet gir seg utslag i gode muslimers adferd overfor det svake kjønn. Islam bygger hele sin eksistens på muslimens totale underkastelse for koranens befalinger. I følge koranen går det ikke an å være « litt » muslim. For at muslimen selv skal akseptere å være Allahs slave og leve i underkastelse og undertrykkelse fra religiøse ledere, gir islam muslimene lov til å sparke nedover, ved å undertrykke, mishandle og voldta sine egne lovlige kvinner og barnebruder. En muslim har lov til å ha fire koner/barnebruder og et ubegrenset antall sexslaver¹. Konseptet er genialt. Når de bedrestilte muslimer har fire hustruer, medfører det at tre mindre bemidlede muslimer

Alle disse tekstene og plakatene fremstiller muslimene som en uensartet gruppe. Gruppen muslimer holdes kollektivt ansvarlige for vold og terror og for å drive lyssky virksomhet. Det er med andre ord mange usanne påstander og historier om muslimer som fremmes i disse tekstene og plakatene hvis innhold trolig spres via leserne. Disse usanne påstandene og historiene om muslimer er godt egnet til å forsterke eller vekke lesernes negative følelser – følelser av frykt, mistenksomhet og sinne. De er også godt egnet til å svekke muslimenes rykte og anseelse i samfunnet, da de rakker ned på muslimene. Slike tekster og plakater kan også inspirere til vold og terror.

6 Konklusjon

I denne rapporten har vi presentert en studie hvor vi har kartlagt innholdet i og omfanget av ytringer av hatefull og diskriminerende karakter mot muslimer. Studien er gjort på oppdrag fra Barne-, ungdoms- og familiedirektoratet. For å kartlegge innholdet i disse ytringene gjorde vi først en kunnskapsoppsummering. Deretter har vi brukt kvantitative analysemetoder basert på egne stordata og maskinlæring for å kartlegge omfanget av og for å gi mer kunnskap om innholdet i ytringer av hatefull eller diskriminerende karakter mot muslimer. Til slutt gjorde vi noen kvalitative tekstanalyser for å gi noen konkrete eksempler på såkalte sammensatte tekster som kan beskrives som hatefulle og diskriminerende.

Kunnskapsoppsummeringen viser at mesteparten av forskningen er fra andre land, og at det finnes begrenset med nordisk litteratur på temaet.

Våre stordata er innhentet fra perioden 2010–2021, og disse viser at:

- det er en markant økning av hatefulle ytringer mot muslimer på norske sosiale medier
- det er en svært liten gruppe som står bak de fleste av disse ytringene.
- muslimer og islam anses som trusler mot Norge, Europa og vestlige verdier

Trekk som går igjen i kunnskapsoppsummeringen vi har gjort, er fremstillingen av muslimer som iboende voldelige og farlige, en tett kobling mellom muslimer generelt og terrorisme, en forestilling om at muslimer er i ferd med å ta over Europa, kritikk av den politiske venstresidas manglende vilje til å ta inn over seg trusselen fra islam, og fremstilling av kulturelle og verdimeslige forskjeller mellom «islam» og «Vesten» som grunnleggende og uoverstigelige. Vi finner disse trekkene også i våre kvantitative og kvalitative analyser. Når vi sammenstiller kunnskapsoppsummeringen med funn fra våre analyser, ser vi med andre ord at det som finnes av hatefulle og diskriminerende ytringer mot muslimer på norske sosiale medier i stor grad samsvarer med funn fra forskning i andre deler av verden. Det er det samme innholdet som ser ut til å gå igjen. Samtalene preget av hatefulle og diskriminerende ytringer mot muslimer på norske sosiale medier, kan således fortolkes som del av en større europeisk samtale der muslimene svertes og rakkes ned på.

Selv om denne rapporten analyserer en stor datamengde, representerer den bare en liten andel av kommentarer fra sosiale medier. I tillegg finnes det i dag flere plattformer som bruker bilder eller videoer som hovedmåte å kommunisere på. Dette er en virkelighet som skaper nye utfordringer for innholdsanalyser. De siste årene har det vært en betydelig fremgang i automatiske metoder for å analysere bilder og videoer ved å bruke maskinlæring, men det er fremdeles teknisk vanskelig. Ikke minst er det problematisk å få tilgang til den type materiale. Både personvern og plattformene selv utgjør til sammen et stort hinder for studier som vil avdekke spredning av hatefulle ytringer i andre typer sosiale medier. Denne rapporten viser likevel hvordan man kan benytte den raske utviklingen innen metoder knyttet til stordata og maskinlæring for å gjøre bedre anslag av omfanget av hatefulle ytringer mot muslimer i sosiale medier. Ved å gjøre kodene våre tilgjengelige, håper vi

at andre forskere kan ta denne analysen videre. Modellene kan enten forbedres eller brukes som referansepunkt for nye algoritmer. Vi åpner også for andre typer studier. For eksempel er det mulig å utvikle et system for å samle inn kommentarer i sanntid. Det betyr at vi kan få et bedre anslag av hatefulle kommentarer ved å hente dem inn før de blir sensurerte eller redigerte. En utvidelse av et annotert datasett kan også muliggjøre en klassifiseringsmodell som er mer raffinert og som er i stand til å identifisere ulike typer hatefulle kommentarer. En mer fokusert analyse av innhold i kommentarene og nyhetssakene kan også gi oss mer innsikt om hva triggerne er for en spesifikk økning i antall hatefulle kommentarer. Da vil vi også kunne undersøke den lille gruppen av hyperaktive profiler ved hjelp av kvalitative metoder.

Referanser

- Aguilera-Carnerero, C. & Azeez, A. H. (2016). 'Islamonausea, not islamophobia': The many faces of cyber hate speech. *Journal of Arab & Muslim Media Research*, 9(1), 21–40.
- Antirasistisk senter. (2018). *Årsrapport*.
https://antirasistisk.no/wp-content/uploads/2020/01/Antirasistisk_årsrapport_2018.pdf
- Awan, I. (2014). Islamophobia and Twitter: A typology of online hate against Muslims on social media. *Policy & Internet*, 6(2), 133–150.
- Awan, I. (2016). Islamophobia on social media: A qualitative analysis of the Facebook's walls of hate. *International Journal of Cyber Criminology*, 10(1), 1–20.
- Awan, I. & Zempi, I. (2015). *We fear for our lives. Offline and online experiences of anti-muslim hostility*. <https://tellmamauk.org/wp-content/uploads/2017/05/We-Fear-For-Our-Lives.pdf>
- Awan, I. & Zempi, I. (2016). The affinity between online and offline anti-Muslim hate crime: Dynamics and impacts. *Aggression & Violent Behavior*, 27, 1–8.
- Awan, I. & Zempi, I. (2017). 'I will blow your face off': Virtual and physical world anti-muslim hate crime. *British Journal of Criminology*, 57(2), 362–380.
- Bai, A., Hammer, H., Yazidi, A. & Engelstad, P. (2014). Constructing sentiment lexicons in Norwegian from a large text corpus. *IEEE 17th International Conference on Computational Science and Engineering*. 231–237.
- Baider, F., Millar, S. & Assimakopoulos, S. (2020). Introduction: Defining, performing and countering hate speech. *Pragmatics and Society* 11(2), 171–176.
- Baldauf, J., Ebner, J. & Guhl, J. (2019). *Hate Speech and Radicalisation Online: The OCCI Research Report*. ISD: London Washington DC Beirut Toronto.
<https://www.isdglobal.org/wp-content/uploads/2019/06/ISD-Hate-Speech-and-Radicalisation-Online-English-Draft-2.pdf>
- Bangstad, S. (2014) *Anders Breivik and the rise of Islamophobia*, London: Zed Books.
- Barne- og likestillingsdepartementet. (2016). *Regjeringens strategi mot hatefulle ytringer 2016–2020*.
https://www.regjeringen.no/contentassets/72293ca5195642249029bf6905ff08be/hatefulleytringer_uu.pdf
- Baumgarten, N., Bick, E., Geyer, K., Iversen, D. A., Kleene, A., Lindø, A. V., ... Nedenskov Petersen, E. (2019). Towards balance and boundaries in public discourse: expressing and perceiving online hate speech (XPEROHS). *RASK: International Journal of Language and Communication*, 50, 1-21.
- Ben-David, A. & Fernández, A. M. (2016). Hate speech and covert discrimination on social media: Monitoring the Facebook pages of extreme-right political parties in Spain. *International Journal of Communication*, 1167–1193.

- Brekke, J. P., Beyer, A. & Enjolras, B. (2019). *Antisemittisme på nett og i sosiale medier i Norge: Kjennetegn, avsendere og motvirkning*. Institutt for Samfunnsforskning. <http://hdl.handle.net/11250/2589085>
- Brindle, A. (2016). Cancer has nothing on Islam: a study of discourses by group elite and supporters of the English defence league. *Critical Discourse Studies*, 13(4), 444–459.
- Burke, S., Diba, P. & Antonopoulos, G. A. (2020). 'You sick, twisted messes': The use of argument and reasoning in Islamophobic and anti-Semitic discussions on Facebook. *Discourse & Society*, 31(4), 374–389.
- Burnap, P. & Williams, M. L. (2015). Cyber hate speech on Twitter: An application of machine classification and statistical modeling for policy and decision making. *Policy & Internet*, 7(2), 223–242.
- Calderón, C. A., de la Vega, G. & Herrero, D. B. (2020). Topic modeling and characterization of hate speech against immigrants on Twitter around the emergence of a far-right party in Spain. *Social Sciences*, 9(11), 1–19. <https://doi.org/10.3390/socsci9110188>
- Carlson, C. R. & Rousselle, H. (2020). Report and repeat: Investigating Facebook's hate speech removal process. *First Monday*, 25(2), <https://doi.org/10.5210/fm.v25i2.10288>
- Cleland, J., Anderson, C. & Aldridge-Deacon, J. (2018). Islamophobia, war and non-Muslims as victims: An analysis of online discourse on an English Defence League message board. *Ethnic and Racial Studies*, 41(9), 1541–1557.
- Dechesne, M., Nilsen, A. B., Paton, N., Sakellariou, A., Helm, G., Baky, A., ... Monsell, L. (2020). *Policy brief. Drivers of self-radicalisation and digital sociability*. <http://www.dare-h2020.org/self-radicalisation-and-digital-sociability.html>
- Downing, J. & Dron, R. (2020). Tweeting Grenfell: Discourse and networks in critical constructions of British Muslim social boundaries on social media. *New Media & Society*, 22(3), 449–469.
- Dvergdsdal, H. (2019). Nevralt nettverk. *Store Norske Leksikon*. https://snl.no/nevralt_nettnettverk
- ECRI. (2016). *ECRI general policy recommendation no. 15 on combatting hate speech*. Strasbourg. <http://rm.coe.int/ecri-general-policy-recommendation-no-15-on-combating-hate-speech/16808b5b01>
- Enggebø, H. & Stubberud, E. (2016). *Hatefulle ytringer. Delrapport 2: Forskning på hat og diskriminering*. Rapport 2016:15. Oslo: Institutt for samfunnsforskning. http://www.kun.no/uploads/7/2/2/3/72237499/delrapport_2_hat_og_diskriminering.pdf
- Ekman, M. (2015). Online Islamophobia and the politics of fear: manufacturing the green scare. *Ethnic and Racial Studies*, 38(11), 1986–2002.
- EISherief, M., Kulkarni, V., Nguyen, D., Wang, W. Y. & Belding, E. (2018). Hate lingo: A target-based linguistic analysis of hate speech in social media. *Proceedings of the International AAAI Conference on Web and Social Media*, 12(1), 1–10.
- Enarsson, T. & Lindgren, S. (2019). Free speech or hate speech? A legal analysis of the discourse about Roma on Twitter. *Information and Communications Technology Law*, 28(1), 1–18.
- Evolvi, G. (2018). Hate in a Tweet: Exploring Internet-Based Islamophobic Discourses. *Religions*, 9(10), 307. <https://doi.org/10.3390/rel9100307>
- Evolvi, G. (2019). #Islamexit: inter-group antagonism on Twitter. *Information, Communication & Society*, 22(3), 386–401.
- Farkas, J., Schou, J. & Neumayer, C. (2018a). Cloaked Facebook pages: Exploring fake Islamist propaganda in social media. *New Media & Society*, 20(5), 1850–1867.
- Farkas, J., Schou, J. & Neumayer, C. (2018b). Platformed antagonism: Racist discourses on fake Muslim Facebook pages. *Critical Discourse Studies*, 15(5), 463–480.
- Fladmoe, A., Nadim, M. & Birkvand, S. R. (2019). *Erfaringer med hatytringer og hets blant LHBT-personer, andre minoritetsgrupper og den øvrige befolkningen*.

- <https://samfunnsforskning.brage.unit.no/samfunnsforskning-xmlui/bitstream/handle/11250/2584665/Erfaringer%2bmed%2bhatytringer.pdf?sequence=2&isAllowed=y>
- Fladmoe, A. & Nadim, M. (2016). *Hatefulle ytringer. Delrapport 1: Forskning på art og omfang*. Oslo: Institutt for samfunnsforskning.
- Fladmoe, A., Nadim, M. & Wessel-Aas, J. (2016). *Hatefulle ytringer. Delrapport 3: Grenseoppgangen mellom ytringsfrihet og strafferettslig vern mot hatefulle ytringer*. Rapport 2016:16, Oslo: Institutt for samfunnsforskning.
https://samfunnsforskning.brage.unit.no/samfunnsforskning-xmlui/bitstream/handle/11250/2442448/Rapport_16_16_HatYtr_Del3_endelig.pdf
- Gaudette, T., Scrivens, R., Davies, G. & Frank, R. (2020). Upvoting extremism: Collective identity formation and the extreme right on Reddit. *New Media & Society*, 1–18.
- Grimmer, J. & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3), 267–297.
- Hanzelka, J. & Schmidt, I. (2017). Dynamics of cyber hate in social media: A comparative analysis of anti-muslim movements in the Czech Republic and Germany. *International Journal of Cyber Criminology*, 11(1), 143–160.
- Hassan, G., Brouillette-Alerie, S., Seraphin, A., Frau-Meigs, D., Lavoie, L., Fetiu, A., Varela, W., Borokhovski, E., Venkatesh, V., Rousseau, C. & Stijn, S. (2018). Exposure to extremist online content could lead to violent radicalization: a systematic review of empirical evidence. *International Journal of Developmental Science*, 12(1–2): 71–88.
<https://doi.org/10.3233/DEV190276>
- Helgesen, Ø., Glavee-Geo, R. Mustafa, G., Nettet, E. & Rice, P. (Red.). (2019). Modeller. *Modeller – Fjordantologien 2019*, 15–37. Oslo: Universitetsforlaget.
- Hoffmann, C. & Moe, V. (Red.). (2017). *Holdninger til jøder og muslimer i Norge 2017 – befolkningsundersøkelse og minoritetsstudie*. Senter for studier av holocaust og livssynsminoriteter. [https://www.hlsenteret.no/forskning/jodisk-historie-og-antisemittisme/befolkningsundersokelse:-holdninger-til-joder-og-a/hl-rapport_29mai-web-\(2\).pdf](https://www.hlsenteret.no/forskning/jodisk-historie-og-antisemittisme/befolkningsundersokelse:-holdninger-til-joder-og-a/hl-rapport_29mai-web-(2).pdf)
- Hervik, P. (2019). Ritualized Opposition in Danish Online Practices of Extremist Language and Thought. *International Journal of Communication*, 13, 3104–3121.
- Hognestad, L. I. & Lamark, H. (2017). Flyktingene kommer! Lokale mediers dekning av flyktningskrisen. *Norsk medietidsskrift*, 24(2), 1-19.
- Hrdina, M. (2016). Identity, Activism and Hatred: Hate Speech against Migrants on Facebook in the Czech Republic in 2015. Identita, aktivismus a nenávist: Nenávistné projevy proti migrantům na Facebooku v České republice v roce 2015. *Nase Spolecnost*(1), 38–47.
- Hudson, Richard. 1975. The meaning of questions. *Language*, 51, 1–31.
- Ilie, Cornelia. 1999. Question-response argumentation in talk shows. *Journal of Pragmatics*, 31, 975–999.
- Jackson, S. J., & Foucault Welles, B. (2015). Hijacking #MNYNYPD: Social media dissent and networked counterpublics. *Journal of communication*(6), 932–952.
- Kress, G. (2010). *Multimodality: a social semiotic approach to contemporary communication*. London: Routledge.
- Kunst, J. R., Sam, D. L. & Ulleberg, P. (2013). Perceived islamophobia: Scale development and validation. *International Journal of Intercultural Relations*, 37(2), 225–237.
- Laurent, M. (2020). Project hatemeter: Helping NGOs and Social Science researchers to analyze and prevent anti-Muslim hate speech on social media. *Procedia Computer Science*, 176, 2143–2153.
- LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep Learning. *Nature* 521, 436–444.
<https://doi.org/10.1038/nature14539>

- LDO (2018). *Likestillings- og diskrimineringsombudets rapport: Hatefulle ytringer i offentlig debatt på nett*. https://www.ldo.no/globalassets/ldo_2019/arkiv/publikasjonsarkiv/fb-og-hatytring/ldo-fb_rapporten_2018.pdf
- LDO (2015). *Likestillings- og diskrimineringsombudets rapport: Hatytringer og hatkriminalitet*. https://www.ldo.no/globalassets/ldo_2019/03_ombudet-og-samfunnet/rapporter/hatefulle-ytringer/hatytringer_og_hatkriminalitet_rapport.pdf
- Lenz, C., Lid, S., Lorenzen, G., Nilsen, A.B., Nustad, P. & Rise, E. (2018). *Tiltak mot hatefulle ytringer. Kunnskaps- og tiltaksoversikt*. Senter for studier av holocaust og livssynsminoriteter. https://bibliotek.bufo.no/BUF/101/Tiltak_mot_hatefulle_ytringer_Kunnskaps_og_tiltaksoversikt.pdf
- Lid, S., Liudden, T. & Stærkebye, M. (2020). *Evaluering av regjeringens strategi mot hatefulle ytringer*. <https://oda.oslomet.no/oda-xmlui/handle/20.500.12199/6513>
- Lingiardi, V., Carone, N., Semeraro, G., Musto, C., D'Amico, M. & Brena, S. (2020). Mapping Twitter hate speech towards social and sexual minorities: a lexicon-based approach to semantic content analysis. *Behaviour & Information Technology*, 39(7), 711–721.
- Magdy, W., Darwish, K. & Abokhodair, N. (2015). *Quantifying public response towards Islam on Twitter after Paris attacks*. arXiv. <https://arxiv.org/abs/1512.04570>
- Mathew, B., Dutt, R., Goyal, P. & Mukherjee, A. (2019a). Spread of hate speech in online social media. *Proceedings of the 10th ACM conference on web science*, 173–182.
- Mathew, B., Saha, P., Tharad, H., Rajgaria, S., Singhanian, P., Kalyan Maitym, S., ... Mukherjee, A. (2019b). Thou shalt not hate: Countering online hate speech. *Proceedings of the International AAAI Conference on Web and Social Media*, 13, 369–380.
- Millar, S., Nielsen, R., Lindø, A. V. & Keyer, K. (2020). The use of hyperlinking as evidential practice in Danish online hate speech. *Pragmatics and Society*, 11(2), 241–261.
- Miller, C. (2017). Australia's anti-Islam right in their own words. Text as data analysis of social media content. *Australian Journal of Political Science*, 52(3), 383–401.
- Miller, C., Arconstanzo, F., Smith, J., Krasodonski-Jones, A., Wiedlitzka, S., Jamali, R. & Dale, J. (2016a). *From Brussels to Brexit: islamophobia, xenophobia, racism and reports of hateful incidents on Twitter*. <http://www.demos.co.uk/wp-content/uploads/2016/07/From-Brussels-to-Brexit-Islamophobia-Xenophobia-Racism-and-Reports-of-Hateful-Incidents-on-Twitter-Research-Prepared-for-Channel-4-Dispatches-%E2%80%98Racist-Britain%E2%80%99-.pdf>
- Miller, C., Smith, J. & Dale, J. (2016b). *Islamophobia on Twitter: March to July 2016*. <http://www.demos.co.uk/wp-content/uploads/2016/08/Islamophobia-on-Twitter-March-to-July-2016-.pdf>
- Mondon, A. & Winter, A. (2017). Articulations of Islamophobia: from the extreme to the mainstream? *Ethnic and Racial Studies*, 40(13), 2151–2179.
- Nadim, M., Thorbjørnsrud, K. og Fladmoe, A. (2021). *Gråsoner og grenseoverskridelser på nettet. En studie av deltagere i opphetede og aggressive nettdebatter*. Oslo: Institutt for samfunnsforskning. <https://samfunnsforskning.brage.unit.no/samfunnsforskning-xmlui/bitstream/handle/11250/2724997/Gr%25C3%25A5soner%20og%20grenseoverskridelser%20p%25C3%25A5%20nettet.pdf?sequence=1&isAllowed=y>
- Nadim, M., Fladmoe, A. & Enjolras, B. (2018). *Måling av omfang av hatefulle ytringer. Metodiske muligheter og utfordringer*. Oslo: Institutt for samfunnsforskning. https://samfunnsforskning.brage.unit.no/samfunnsforskning-xmlui/bitstream/handle/11250/2478635/Rapport_1-18_web.pdf?sequence=1&isAllowed=y
- Nadim, M., Fladmoe, A. & Wessel-Aas, J. (2016). *Hatefulle ytringer på internett. Omfang, forebygging og juridiske grenser*. Oslo: Institutt for samfunnsforskning.

- Nadim, M., Fladmoe, A. & Wessel-Aas, J. (2017). *Hatefulle ytringer på internett. Omfang, forebygging og juridiske grenser*. Oslo: Institutt for samfunnsforskning.
<https://samfunnsforskning.brage.unit.no/samfunnsforskning-xmlui/handle/11250/2442451>
- Nilsen, A.B. (2014a). Hatets visuelle og verbale retorikk. I S. Østerud (Red.), *22. Juli – forstå, forklare, forebygge*. Oslo: Abstrakt forlag, 199–222.
- Nilsen, A. B. (2014b). *Hatprat*. Oslo: Cappelen Damm akademisk.
- Nilsen, A. B., Paton, N., Dechesne, M., Sakellariou, A., Helm, G., Baky, A. & Monsell, L. (2020). *Cross-National Level Report on Digital Sociability and drivers of self-radicalisation in Europe*. <http://www.dare-h2020.org/self-radicalisation-and-digital-sociability.html>
- Nilsen, A. B. & Paton, N. (2020). Primary drivers of self-radicalisation – Norway. *Prosjektrapport DARE – Dialogue about Radicalisation and Equality*.
http://www.dareh2020.org/uploads/1/2/1/7/12176018/5.1_norway1.pdf
- Nilsen, A.B. (2022). *Elevers språkbruk. Hvordan forbedre relasjoner i klasserommet*. Cappelen Damm Akademisk
- Oboler, A. (2013). *Islamophobia on the internet. The growth of online hate targeting muslims*.
https://www.researchgate.net/publication/271706783_Islamophobia_on_the_Internet_The_growth_of_online_hate_targeting_Muslims
- Oboler, A. (2016). The Normalisation of Islamophobia through Social Media: Facebook. I I. Awan (Red.), *Islamophobia in Cyberspace. Hate crime goes viral*. London: Routledge.
- Olteanu, A., Castillo, C., Boy, J. & Varshney, K. (2018). The effect of extremist violence on hateful speech online. *Proceedings of the International AAAI Conference on Web and Social Media*, 12(1).
- Oslo politidistrikt. (2020). *Hatkriminalitet. Anmeldt hatkriminalitet 2020*.
https://www.politiet.no/globalassets/dokumenter/oslo/rapporter/anmeldt-hatkriminalitet-oslo/anmeldt-hatkriminalitet-i-oslo-2020.pdf? t_id=dn-herAKQBySkIwA40kkRg%3d%3d& t uuid=cMDqh8spRa6tQ87dnHX6IA& t q=hatkriminalitet& t tags=language%3ano%2csiteid%3ab78013b9-3a34-4dad-8a4f-cc70168f7f39%2candquerymatch& t hit.id=Politiet Web Models Media GenericMedia/629c7d8f-f4cb-40ae-b5f9-e19f99b30aa3& t hit.pos=5
- Otoni, R., Cunha, E., Magno, G., Bernardina, P., Meira Jr, W. & Almeida, V. (2018). Analyzing right-wing YouTube channels: Hate, violence and discrimination. *Proceedings of the 10th ACM Conference on Web Science*, 323–332.
- Piazza, J. A. (2020). Politician hate speech and domestic terrorism. *International Interactions*, 46:3, 431–453. <https://doi.org/10.1080/03050629.2020.1739033>
- Poletto, F., Stranisci, M., Sanguinetti, M., Patti, V. & Bosco, C. (2017). Hate speech annotation: Analysis of an Italian Twitter corpus. In *4th Italian Conference on Computational Linguistics, 2006*, 1–6.
- Poole, E., Giraud, E. H. & de Quincey, E. (2019). Contesting #StopIslam: The Dynamics of a counter-narrative against right-wing populism. *Open Library of Humanities*, 5(1).
- Poole, E., Giraud, E. H. & de Quincey, E. (2020). Tactical interventions in online hate speech: The case of #stopIslam. *New Media and Society*, 1–28.
- Ravndal, J.A. (2013). Anders Behring Breiviks bruk av internett og sosiale medier. I I. M. Sunde (Red.), *Forebygging av radikaliserings og voldelig ekstremisme på internett*.
- Regjeringen. (2017). *Regjeringens handlingsplan mot radikaliserings og voldelig ekstremisme*.
<https://www.regjeringen.no/no/sub/radikaliserings/aktuelt/nye-tiltak-i-handlingsplanen-mot-radikaliserings-og-voldelig-ekstremisme/id2542460/>
- Rudin, C. (2019). Stop explaining black box machine learning for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215.
- Scharwächter, E. & Müller, E. (2020). Does terrorism trigger online hate speech? On the association of events and time series. *Annals of Applied Statistics*, 14(3), 1285–1303.

- Schmidt-Radefeldt, Jürgen. 1977. On so-called «rhetorical» questions. *Journal of Pragmatics*, 1 (4), 375–392.
- Siapera, E. (2019). Organised and Ambient Digital Racism: Multidirectional Flows in the Irish Digital Sphere. *Open Library of Humanities*, 5(1), 1–34.
- Soral, W., Liu, J. & Bilewicz, M. (2020). Media of contempt: Social media consumption predicts normative acceptance of anti-Muslim hate speech and islamoprejudice. *International Journal of Conflict and Violence*, 14(1), 1–13.
- Staksrud, E., Steen-Johnsen, K., Enjolras, B., Gustafsson, M. H., Ihlebæk, K. A., Midtbøen, A. H., ... Utheim, M. (2014). *Status for ytringsfriheten i Norge. Resultater fra befolkningsundersøkelsen 2014*.
https://samfunnsforskning.brage.unit.no/samfunnsforskning-xmlui/bitstream/handle/11250/2389658/Rev_Ytringsfrihet%20i%20Norge%20Holdninger%20og%20erfaringer%20rev2.pdf?sequence=4
- Straffeloven. (2005). *Lov om straff* (LOV-2005-05-20-28). Lovdata. <https://lovdata.no/lov/2005-05-20-28>
- Strømmen, Ø. (2014). *I hatets fotspor*. Oslo: Cappelen Damm
- Tell MAMA. (2018). *Beyond the incidence. Outcomes for victims of anti-muslim prejudice. Tell MAMA annual report 2017*. <https://tellmamauk.org/wp-content/uploads/2018/07/Tell-MAMA-Report-2017.pdf>
- Tell MAMA. (2019). *The impact of the Christchurch terror attack. Tell MAMA interim report 2019*. <https://www.tellmamauk.org/wp-content/uploads/2020/03/The-Impact-of-the-ChristChurch-Attack-Tell-MAMA-Interim-Report-2019-PP.pdf>
- Veledar, A. & Burkal, R. (2018). *Hatefulle ytringer i offentlig debatt og på nett*. https://www.ido.no/globalassets/ido_2019/arkiv/publikasjonsarkiv/fb-og-hatytring/ido-fb-rapporten_2018.pdf
- Vidgen, B. & Yasseri, T. (2020). Detecting weak and strong Islamophobic hate speech on social media. *Journal of Information Technology & Politics*, 17(1), 66–78.
- Vidgen, B., Yasseri, T. & Margetts, H. (2021). Islamophobes are not all the same! A study of far right actors on Twitter. *Journal of Policing, Intelligence and Counter Terrorism*, 1–23.
- Waldron, J. (2012). *The Harm in Hate Speech*. Cambridge: Harvard University Press
- Waseem, Z. & Hovy, D. (2016). Hateful symbols or hateful people? Predictive features for hate speech detection on Twitter. *Proceedings of the NAACL student research workshop*, 88–93.
- Wasvik, M. (2017). *Vi vil ikke leke med deg fordi du er brun – en undersøkelse av opplevd rasisme blant ungdom*. Oslo, 2017. <http://antirasistisk.no/wp-content/uploads/2017/08/Vi-vil-ikke-leke-med-deg-fordi-du-er-brun-en-unders%C3%B8kelse-av-opplevd-rasisme-blant-ungdom.pdf>
- Williams, M. L., Burnap, P., Javed, A., Liu, H. & Ozalp, S. (2020). Hate in the Machine: Anti-Black and Anti-Muslim Social Media Posts as Predictors of Offline Racially and Religiously Aggravated Crime. *British Journal of Criminology*, 60(1), 93–117.
- Zhang, Z. & Luo, L. (2019). Hate speech detection: A solved problem? The challenging case of long tail on Twitter. *Semantic Web*, 10(5), 925–945.
- Zhang, Z., Robinson, D. & Tepper, J. (2018). Detecting Hate Speech on Twitter using a Convolution-GRU Based Deep Neural Network. *European Semantic Web Conference*. 745–760.
- Zuleta, L., Burkal, R. & Thomsen, T. B. (2017). *Hadefulle ytringer i den offentlige online debat*. https://menneskeret.dk/sites/menneskeret.dk/files/media/dokumenter/udgivelser/liigebehandling_2017/rapport_om_hadefulde_ytringer_2._oplag_2017.pdf

Øistad, B. S. & Veledar, A. (2021). *Hatefulle ytringer på nett*.

https://www.ldo.no/globalassets/ldo_2019/03_ombudet-og-samfunnet/rapporter/hatefulle-ytringer/ldo_hatefulle_ytringer_pa_nett.pdf

Øvrebø, S. B. (2018). Rettslig bedømmelse av hatefulle ytringer: En vurdering av norsk høyesterettspraksis i lys av amerikansk rett. [Masteroppgave, Universitetet i Oslo]. DUO. <http://urn.nb.no/URN:NBN:no-70167>

Vedlegg - Søkehistorikk

Problemstilling, avgrensninger

Problemstilling: Kunnskapsoppsummering av eksisterende kunnskap om hatytringer mot muslimer, fremsatt på digitale plattformer. Dette søket baserer seg på et tidligere litteratursøk utført i 2017. Denne nye, utvidede versjonen er utviklet i samarbeid med oppdragsgiver og ut fra oppdragsgivers behov.

Søkets hovedelementer: Hatefulle ytringer - muslimer - nettbasert

Utstrakt bruk av søk på fraser i søket er brukt for å identifisere mest mulig relevante treff og øke søkets presisjon. Et bredere søk kunne vært satt opp med bruk av nærhetsoperatører, men ut fra søkets natur ble det vurdert som ikke nødvendig nå. Det er altså trolig at noen artikler som potensielt kunne være relevante ikke er identifisert.

Det er ikke søkt på religion generelt. Dette kan ha introdusert en svakhet i søket ettersom noen forfattere kanskje vil trekke frem dette elementet i tittel/sammendrag, men først konkretisere for eksempel islam i fullteksten.

Spesielt i nyere forskning kan man tenke at hatefulle ytringer i større grad fremmes digitalt. Dette elementet kunne derfor vært utelatt for forskning fra senere år for å øke sensitiviteten i søket, men det vil naturlig nok identifisere en del forskning som ikke er relevant for problemstillingen. Det har derfor ikke blitt tatt med i dette søket – nettopp med det formålet å holde presisjonen høy

Basert på oppdragsgivers behov og problemstilling er søket avgrenset der databasene tillot det og ut fra totalt antall treff. Dette gjelder publikasjonstype: fagfelleverderte tidsskrift (conference proceedings er tatt med der hvor de var del av trefflisten), tidsavgrensning: publisert i perioden 2010 – 2020, og språk: norsk/skandinavisk, engelsk.

Valg av databaser

Valg av databaser ble gjort på bakgrunn av problemstilling, erfaringer med tidligere søk om Hatefulle ytringer (2017), og hvilke databaser som er tilgjengelige ved HiOA. Valget falt på følgende internasjonale, vitenskapelige databaser: Education Source, SocINDEX, Academic Search Ultimate, Web of Science, og Scopus.

Gjennom dette ble et bredt sett faglige perspektiv representert i litteratursøket. For nærmere beskrivelse av hver enkelt database, se <https://bibsyst-almaprmo.hosted.exlibrisgroup.com/primo-explore/dbsearch?vid=HIOA>

Det er gjennomført et strukturert søk, hvor det er søkt etter søkeordene i tittel, sammendrag eller nøkkelord. I de databasene som har kontrollerte emneord, er det også søkt på disse.

Emneordsregistrene (*thesaurus*) er spesifikke for den enkelte database. Derfor er disse delene av søkene forskjellige, mens søkene forøvrig er mest mulig like på tvers av basene.

Skandinaviske databaser:

Det er langt bedre søkefunksjonalitet og muligheter til avanserte, strukturerte søk i de større, vitenskapelige databasene, enn i de skandinaviske. Dette må prege fremgangsmåten i litteratursøket, som justeres individuelt, både med tanke på teknisk søkefunksjonalitet og innhold i databasene.

Generelt er det til dels svært krevende å gjennomføre større, flerdelte søk (hatefulle ytringer og muslimer og nettbasert) i de skandinaviske databasene, samtidig som de ikke nødvendigvis er tilrettelagt for enkelt å kunne eksportere treffene til EndNote. Det er derfor viktig at å fokusere på de mest pregnante søketermene i disse søkene.

Med bakgrunn i erfaringene fra søk utført til prosjektet Hatefulle ytringer (2017-søket), er det søkt i nordiske databaser som antas å gi flest relevante treff, ut fra muligheter i grensesnittet. Det er brukt en forenklet søkestreng, da leddet «muslim* OR islam*» til dels ikke er mulig å legge til, delvis vil medføre en betydelig arbeidsmengde med hensyn til å eksportere et begrenset antall treff til Endnote – og delvis ikke er mulig pga. begrensninger i hvor mange ledd de ulike databasene kan håndtere. Man kan velge å gjennomføre søkene, og så screene gjennom trefflistene før man eventuelt velger å eksportere unike, relevante treff.

Det er gjennomført søk i Swepub, Diva, Den danske forskningsdatabase, Oria samt Google Scholar for å fange opp nordiske referanser som ikke omfattes av de internasjonale databasene. Oria håndterer ikke lange søkestrenger, derfor må antall søketermer begrenses relativt kraftig. Dette gjelder jo da særlig den delen av søket som spesifiserer hatefulle ytringer. Oria håndterer heller ikke trunkering av så mange termer, så parallelt med å redusere antall søkeord må man også vurdere hvor det er nødvendig å søke på både entalls- og flertallsformer.

For å fange opp mer av ikke-engelskspråklige publikasjoner og grått materiale er det søkt i Google Scholar (norsk, svensk og dansk søkemotor). Disse søkene er gjort ved hjelp av nærhetsoperator (AROUND) – som kobler to ledd: netthat og islam, og avgrensa til 2010 – 2020. Dette gir langt færre, men presumptivt mer relevante treff enn åpne søk på disse to leddene. Ved å sortere treff på relevans kan man relativt fort screene gjennom og se etter relevante treff, som så (enkeltvis) kan eksporteres til endnote for dublettkontroll.

Det er også søkt i databasen Open Grey etter grå litteratur. Her er referanselista knyttet mot Zotero, ikke endnote, så treffene er ikke eksportert. Det er relativt få treff, som kan screenes før man velger om de skal inkluderes.

Database: Education Source
Dato: 09.11.2020
Treff: 5

#	Query	Limiters/Expanders	Results
S1	DE "Libel & slander"	Search modes - Boolean/Phrase	1,006
S2	TI (cyberhate OR "cyber bullying" OR cyberbullying OR "hate speech" OR "hate rhetoric" OR "hate utterance" OR "hate utterances" OR "hateful speech" OR "hateful rhetoric" OR "hateful utterance" OR "hateful utterances" OR "degrading speech" OR "degrading rhetoric" OR "degrading utterance" OR "degrading utterances" OR "threatening speech" OR "threatening rhetoric" OR "threatening utterance" OR "threatening utterances" OR "harassing speech" OR "harassing rhetoric" OR "harassing utterance" OR "harassing utterances" OR "stigmatizing speech" OR "stigmatizing rhetoric" OR "stigmatizing utterance" OR "stigmatizing utterances" OR "vile speech" OR "vile rhetoric" OR "vile utterance" OR "vile utterances" OR "offensive speech" OR "offensive rhetoric" OR "offensive utterance" OR "offensive utterances" OR "outrageous speech" OR "outrageous rhetoric" OR "outrageous utterance" OR "outrageous utterances" OR "discriminatory speech" OR "discriminatory rhetoric" OR "discriminatory utterance" OR "discriminatory utterances" OR "discriminating speech" OR "discriminating rhetoric" OR "discriminating utterance" OR "discriminating utterances" OR "harmful speech" OR "harmful rhetoric" OR "harmful utterance" OR "harmful utterances" OR "racist speech" OR "racist rhetoric" OR "racist utterance" OR "racist utterances" OR "discriminative speech" OR "discriminative rhetoric" OR "discriminative utterance" OR "discriminative utterances" OR "derogatory speech" OR "derogatory rhetoric" OR "derogatory utterance" OR "derogatory utterances" OR "extreme speech" OR "extreme rhetoric" OR "extreme utterance" OR "extreme utterances" OR "prejudice speech" OR "prejudice rhetoric" OR "prejudice utterance" OR "prejudice utterances" OR "prejudicial speech" OR "prejudicial rhetoric" OR "prejudicial utterance" OR "prejudicial utterances" OR "intolerant speech" OR "intolerant rhetoric" OR "intolerant utterance" OR "intolerant utterances" OR "excitable speech" OR "excitable rhetoric" OR "excitable utterance" OR "excitable utterances" OR "excitable rhetoric" OR "extremist speech" OR "extremist rhetoric" OR "extremist utterance" OR "extremist utterances" OR "online hate" OR "rhetoric of hate" OR "rhetoric of discrimination" OR "harm speech") OR AB (cyberhate OR "cyber bullying" OR cyberbullying OR "hate speech" OR "hate rhetoric" OR "hate utterance" OR "hate utterances" OR "hateful speech" OR "hateful rhetoric" OR "hateful utterance" OR "hateful utterances" OR "degrading speech" OR "degrading rhetoric" OR "degrading utterance" OR "degrading utterances" OR "threatening speech" OR "threatening rhetoric" OR "threatening utterance" OR "threatening utterances" OR "harassing speech" OR "harassing rhetoric" OR "harassing utterance" OR "harassing utterances" OR "stigmatizing speech" OR "stigmatizing rhetoric" OR "stigmatizing	Search modes - Boolean/Phrase	2,432

	<p>utterance" OR "stigmatizing utterances" OR "vile speech" OR "vile rhetoric" OR "vile utterance" OR "vile utterances" OR "offensive speech" OR "offensive rhetoric" OR "offensive utterance" OR "offensive utterances" OR "outrageous speech" OR "outrageous rhetoric" OR "outrageous utterance" OR "outrageous utterances" OR "discriminatory speech" OR "discriminatory rhetoric" OR "discriminatory utterance" OR "discriminatory utterances" OR "discriminating speech" OR "discriminating rhetoric" OR "discriminating utterance" OR "discriminating utterances" OR "harmful speech" OR "harmful rhetoric" OR "harmful utterance" OR "harmful utterances" OR "racist speech" OR "racist rhetoric" OR "racist utterance" OR "racist utterances" OR "discriminative speech" OR "discriminative rhetoric" OR "discriminative utterance" OR "discriminative utterances" OR "derogatory speech" OR "derogatory rhetoric" OR "derogatory utterance" OR "derogatory utterances" OR "extreme speech" OR "extreme rhetoric" OR "extreme utterance" OR "extreme utterances" OR "prejudice speech" OR "prejudice rhetoric" OR "prejudice utterance" OR "prejudice utterances" OR "prejudicial speech" OR "prejudicial rhetoric" OR "prejudicial utterance" OR "prejudicial utterances" OR "intolerant speech" OR "intolerant rhetoric" OR "intolerant utterance" OR "intolerant utterances" OR "excitable speech" OR "excitable rhetoric" OR "excitable utterance" OR "excitable utterances" OR "excitable rhetoric" OR "extremist speech" OR "extremist rhetoric" OR "extremist utterance" OR "extremist utterances" OR "online hate" OR "rhetoric of hate" OR "rhetoric of discrimination" OR "harm speech") OR KW (cyberhate OR "cyber bullying" OR cyberbullying OR "hate speech" OR "hate rhetoric" OR "hate utterance" OR "hate utterances" OR "hateful speech" OR "hateful rhetoric" OR "hateful utterance" OR "hateful utterances" OR "degrading speech" OR "degrading rhetoric" OR "degrading utterance" OR "degrading utterances" OR "threatening speech" OR "threatening rhetoric" OR "threatening utterance" OR "threatening utterances" OR "harassing speech" OR "harassing rhetoric" OR "harassing utterance" OR "harassing utterances" OR "stigmatizing speech" OR "stigmatizing rhetoric" OR "stigmatizing utterance" OR "stigmatizing utterances" OR "vile speech" OR "vile rhetoric" OR "vile utterance" OR "vile utterances" OR "offensive speech" OR "offensive rhetoric" OR "offensive utterance" OR "offensive utterances" OR "outrageous speech" OR "outrageous rhetoric" OR "outrageous utterance" OR "outrageous utterances" OR "discriminatory speech" OR "discriminatory rhetoric" OR "discriminatory utterance" OR "discriminatory utterances" OR "discriminating speech" OR "discriminating rhetoric" OR "discriminating utterance" OR "discriminating utterances" OR "harmful speech" OR "harmful rhetoric" OR "harmful utterance" OR "harmful utterances" OR "racist speech" OR "racist rhetoric" OR "racist utterance" OR "racist utterances" OR "discriminative speech" OR "discriminative rhetoric" OR "discriminative utterance" OR "discriminative utterances" OR "derogatory speech" OR "derogatory rhetoric" OR "derogatory utterance" OR "derogatory utterances" OR "extreme speech" OR "extreme rhetoric" OR "extreme utterance" OR "extreme utterances" OR "prejudice speech" OR "prejudice rhetoric" OR "prejudice utterance" OR "prejudice utterances" OR "prejudicial speech" OR "prejudicial rhetoric" OR "prejudicial utterance" OR "prejudicial utterances" OR "intolerant speech" OR "intolerant rhetoric" OR "intolerant utterance" OR "intolerant utterances" OR "excitable speech" OR "excitable rhetoric" OR "excitable utterance" OR "excitable utterances" OR "excitable rhetoric" OR "extremist speech" OR "extremist rhetoric" OR "extremist utterance" OR</p>		
--	--	--	--

	"extremist utterances" OR "online hate" OR "rhetoric of hate" OR "rhetoric of discrimination" OR "harm speech")		
S3	S1 OR S2	Search modes - Boolean/Phrase	3,425
S4	DE "Education of Muslims"	Search modes - Boolean/Phrase	120
S5	TI (muslim* OR islam* OR moslem*) OR AB (muslim* OR islam* OR moslem*) OR KW (muslim* OR islam* OR moslem*)	Search modes - Boolean/Phrase	22,476
S6	S4 OR S5	Search modes - Boolean/Phrase	22,487
S7	DE "Online chat"	Search modes - Boolean/Phrase	1,017
S8	DE "Internet forums"	Search modes - Boolean/Phrase	3,038
S9	DE "Digital media"	Search modes - Boolean/Phrase	3,492
S10	DE "Internet users" OR DE "Social media"	Search modes - Boolean/Phrase	14,262
S11	DE "Backchannels (Social media)"	Search modes - Boolean/Phrase	32
S12	TI (online OR "social media*" OR digital* OR internet* OR Facebook OR blog* OR Twitter OR chat* OR youtube OR WhatsApp* OR Messenger OR "web sites" OR websites OR "social netw*" OR cyber OR cyberspace) OR AB (online OR "social media*" OR digital* OR internet* OR Facebook OR blog* OR Twitter OR chat* OR youtube OR WhatsApp* OR Messenger OR "web sites" OR websites OR "social netw*" OR cyber OR cyberspace) OR KW (online OR "social media*" OR digital* OR internet* OR Facebook OR blog* OR Twitter OR chat* OR youtube OR WhatsApp* OR Messenger OR "web sites" OR websites OR "social netw*" OR cyber OR cyberspace)	Search modes - Boolean/Phrase	289,198
S13	S7 OR S8 OR S9 OR S10 OR S11 OR S12	Search modes - Boolean/Phrase	292,469
S14	S3 AND S6 AND S13	Search modes - Boolean/Phrase	10
S15	S3 AND S6 AND S13	Limiters - Scholarly (Peer Reviewed) Journals Search modes - Boolean/Phrase	5

Database: Academic Source Ultimate
Dato: 09.11 2020
Treff: 39

#	Query	Limiters/Expanders	Results
S1	DE "HATE speech" OR DE "ONLINE hate speech" OR DE "Cyberbullying"	Search modes - Boolean/Phrase	3,419
S2	TI (cyberhate OR "cyber bullying" OR cyberbullying OR "hate speech" OR "hate rhetoric" OR "hate utterance" OR "hate utterances" OR "hateful speech" OR "hateful rhetoric" OR "hateful utterance" OR "hateful utterances" OR "degrading speech" OR "degrading rhetoric" OR "degrading utterance" OR "degrading utterances" OR "threatening speech" OR "threatening rhetoric" OR "threatening utterance" OR "threatening utterances" OR "harassing speech" OR "harassing rhetoric" OR "harassing utterance" OR "harassing utterances" OR "stigmatizing speech" OR "stigmatizing rhetoric" OR "stigmatizing utterance" OR "stigmatizing utterances" OR "vile speech" OR "vile rhetoric" OR "vile utterance" OR "vile utterances" OR "offensive speech" OR "offensive rhetoric" OR "offensive utterance" OR "offensive utterances" OR "outrageous speech" OR "outrageous rhetoric" OR "outrageous utterance" OR "outrageous utterances" OR "discriminatory speech" OR "discriminatory rhetoric" OR "discriminatory utterance" OR "discriminatory utterances" OR "discriminating speech" OR "discriminating rhetoric" OR "discriminating utterance" OR "discriminating utterances" OR "harmful speech" OR "harmful rhetoric" OR "harmful utterance" OR "harmful utterances" OR "racist speech" OR "racist rhetoric" OR "racist utterance" OR "racist utterances" OR "discriminative speech" OR "discriminative rhetoric" OR "discriminative utterance" OR "discriminative utterances" OR "derogatory speech" OR "derogatory rhetoric" OR "derogatory utterance" OR "derogatory utterances" OR "extreme speech" OR "extreme rhetoric" OR "extreme utterance" OR "extreme utterances" OR "prejudice speech" OR "prejudice rhetoric" OR "prejudice utterance" OR "prejudice utterances" OR "prejudicial speech" OR "prejudicial rhetoric" OR "prejudicial utterance" OR "prejudicial utterances" OR "intolerant speech" OR "intolerant rhetoric" OR "intolerant utterance" OR "intolerant utterances" OR "excitable speech" OR "excitable rhetoric" OR "excitable utterance" OR "excitable utterances" OR "excitable rhetoric" OR "extremist speech" OR "extremist rhetoric" OR "extremist utterance" OR "extremist utterances" OR "online hate" OR "rhetoric of hate" OR "rhetoric of discrimination" OR "harm speech")) OR AB (cyberhate OR "cyber bullying" OR cyberbullying OR "hate speech" OR "hate rhetoric" OR "hate utterance" OR "hate utterances" OR "hateful speech" OR "hateful rhetoric" OR "hateful utterance" OR "hateful utterances" OR "degrading speech" OR "degrading rhetoric" OR "degrading utterance" OR "degrading utterances" OR "threatening speech" OR "threatening rhetoric" OR "threatening utterance" OR "threatening utterances" OR "harassing speech" OR "harassing rhetoric" OR "harassing utterance" OR "harassing utterances" OR "stigmatizing speech" OR "stigmatizing rhetoric" OR "stigmatizing utterance" OR "stigmatizing utterances" OR "vile speech" OR "vile rhetoric" OR "vile utterance" OR "vile utterances" OR "offensive speech" OR "offensive rhetoric" OR "offensive utterance" OR "offensive utterances" OR "outrageous speech" OR "outrageous rhetoric" OR "outrageous utterance" OR "outrageous utterances" OR "discriminatory speech" OR "discriminatory rhetoric" OR	Search modes - Boolean/Phrase	4,310

	<p>"discriminatory utterance" OR "discriminatory utterances" OR "discriminating speech" OR "discriminating rhetoric" OR "discriminating utterance" OR "discriminating utterances" OR "harmful speech" OR "harmful rhetoric" OR "harmful utterance" OR "harmful utterances" OR "racist speech" OR "racist rhetoric" OR "racist utterance" OR "racist utterances" OR "discriminative speech" OR "discriminative rhetoric" OR "discriminative utterance" OR "discriminative utterances" OR "derogatory speech" OR "derogatory rhetoric" OR "derogatory utterance" OR "derogatory utterances" OR "extreme speech" OR "extreme rhetoric" OR "extreme utterance" OR "extreme utterances" OR "prejudice speech" OR "prejudice rhetoric" OR "prejudice utterance" OR "prejudice utterances" OR "prejudicial speech" OR "prejudicial rhetoric" OR "prejudicial utterance" OR "prejudicial utterances" OR "intolerant speech" OR "intolerant rhetoric" OR "intolerant utterance" OR "intolerant utterances" OR "excitable speech" OR "excitable rhetoric" OR "excitable utterance" OR "excitable utterances" OR "excitable rhetoric" OR "extremist speech" OR "extremist rhetoric" OR "extremist utterance" OR "extremist utterances" OR "online hate" OR "rhetoric of hate" OR "rhetoric of discrimination" OR "harm speech") OR KW (cyberhate OR "cyber bullying" OR cyberbullying OR "hate speech" OR "hate rhetoric" OR "hate utterance" OR "hate utterances" OR "hateful speech" OR "hateful rhetoric" OR "hateful utterance" OR "hateful utterances" OR "degrading speech" OR "degrading rhetoric" OR "degrading utterance" OR "degrading utterances" OR "threatening speech" OR "threatening rhetoric" OR "threatening utterance" OR "threatening utterances" OR "harassing speech" OR "harassing rhetoric" OR "harassing utterance" OR "harassing utterances" OR "stigmatizing speech" OR "stigmatizing rhetoric" OR "stigmatizing utterance" OR "stigmatizing utterances" OR "vile speech" OR "vile rhetoric" OR "vile utterance" OR "vile utterances" OR "offensive speech" OR "offensive rhetoric" OR "offensive utterance" OR "offensive utterances" OR "outrageous speech" OR "outrageous rhetoric" OR "outrageous utterance" OR "outrageous utterances" OR "discriminatory speech" OR "discriminatory rhetoric" OR "discriminatory utterance" OR "discriminatory utterances" OR "discriminating speech" OR "discriminating rhetoric" OR "discriminating utterance" OR "discriminating utterances" OR "harmful speech" OR "harmful rhetoric" OR "harmful utterance" OR "harmful utterances" OR "racist speech" OR "racist rhetoric" OR "racist utterance" OR "racist utterances" OR "discriminative speech" OR "discriminative rhetoric" OR "discriminative utterance" OR "discriminative utterances" OR "derogatory speech" OR "derogatory rhetoric" OR "derogatory utterance" OR "derogatory utterances" OR "extreme speech" OR "extreme rhetoric" OR "extreme utterance" OR "extreme utterances" OR "prejudice speech" OR "prejudice rhetoric" OR "prejudice utterance" OR "prejudice utterances" OR "prejudicial speech" OR "prejudicial rhetoric" OR "prejudicial utterance" OR "prejudicial utterances" OR "intolerant speech" OR "intolerant rhetoric" OR "intolerant utterance" OR "intolerant utterances" OR "excitable speech" OR "excitable rhetoric" OR "excitable utterance" OR "excitable utterances" OR "excitable rhetoric" OR "extremist speech" OR "extremist rhetoric" OR "extremist utterance" OR "extremist utterances" OR "online hate" OR "rhetoric of hate" OR "rhetoric of discrimination" OR "harm speech"))</p>		
S3	S1 OR S2	Search modes - Boolean/Phrase	5,116

S4	DE "MUSLIMS"	Search modes - Boolean/Phrase	14,646
S5	DE "ISLAMIZATION"	Search modes - Boolean/Phrase	322
S6	S4 OR S5	Search modes - Boolean/Phrase	14,936
S7	TI ((muslim* OR islam* OR moslem*)) OR AB ((muslim* OR islam* OR moslem*)) OR KW ((muslim* OR islam* OR moslem*))	Search modes - Boolean/Phrase	161,227
S8	S6 OR S7	Search modes - Boolean/Phrase	162,070
S9	DE "ONLINE chat"	Search modes - Boolean/Phrase	1,902
S10	DE "INTERNET forums" OR DE "INSTANT messaging" OR DE "INTERNET"	Search modes - Boolean/Phrase	79,565
S11	DE "DIGITAL media"	Search modes - Boolean/Phrase	11,050
S12	DE "SOCIAL media"	Search modes - Boolean/Phrase	30,188
S13	DE "ONLINE comments"	Search modes - Boolean/Phrase	478
S14	S9 OR S10 OR S11 OR S12 OR S13	Search modes - Boolean/Phrase	119,416
S15	TI (online OR "social media*" OR digital* OR internet* OR Facebook OR blog* OR Twitter OR chat* OR youtube OR WhatsApp* OR Messenger OR "web sites" OR websites OR "social netw*" OR cyber OR cyberspace) OR AB (online OR "social media*" OR digital* OR internet* OR Facebook OR blog* OR Twitter OR chat* OR youtube OR WhatsApp* OR Messenger OR "web sites" OR websites OR "social netw*" OR cyber OR cyberspace) OR KW (online OR "social media*" OR digital* OR internet* OR Facebook OR blog* OR Twitter OR chat* OR youtube OR WhatsApp* OR Messenger OR "web sites" OR websites OR "social netw*" OR cyber OR cyberspace)	Search modes - Boolean/Phrase	1,121,561
S16	S14 OR S15	Search modes - Boolean/Phrase	1,136,122
S17	S3 AND S8 AND S16	Search modes - Boolean/Phrase	57
S18	S3 AND S8 AND S16	Limiters - Published Date: 20100101-20201231 Search modes - Boolean/Phrase	53
S19	S3 AND S8 AND S16	Limiters - Scholarly (Peer Reviewed) Journals; Published Date: 20100101-20201231	39

		Search modes - Boolean/Phrase	
--	--	----------------------------------	--

Database: SocIndex
Dato: 09.11 2020
Treff: 17

#	Query	Limiters/Expanders	Results
S1	DE "LIBEL & slander" OR DE "HATE speech"	Search modes - Boolean/Phrase	636
S2	DE "DISCRIMINATORY language"	Search modes - Boolean/Phrase	30
S3	S1 OR S2	Search modes - Boolean/Phrase	664
S4	TI ((cyberhate OR cyberbully* OR "hate speech" OR "hate rhetoric" OR "hate utterance" OR "hate utterances" OR "hateful speech" OR "hateful rhetoric" OR "hateful utterance" OR "hateful utterances" OR "degrading speech" OR "degrading rhetoric" OR "degrading utterance" OR "degrading utterances" OR "threatening speech" OR "threatening rhetoric" OR "threatening utterance" OR "threatening utterances" OR "harassing speech" OR "harassing rhetoric" OR "harassing utterance" OR "harassing utterances" OR "stigmatizing speech" OR "stigmatizing rhetoric" OR "stigmatizing utterance" OR "stigmatizing utterances" OR "vile speech" OR "vile rhetoric" OR "vile utterance" OR "vile utterances" OR "offensive speech" OR "offensive rhetoric" OR "offensive utterance" OR "offensive utterances" OR "outrageous speech" OR "outrageous rhetoric" OR "outrageous utterance" OR "outrageous utterances" OR "discriminatory speech" OR "discriminatory rhetoric" OR "discriminatory utterance" OR "discriminatory utterances" OR "discriminating speech" OR "discriminating rhetoric" OR "discriminating utterance" OR "discriminating utterances" OR "harmful speech" OR "harmful rhetoric" OR "harmful utterance" OR "harmful utterances" OR "racist speech" OR "racist rhetoric" OR "racist utterance" OR "racist utterances" OR "discriminative speech" OR "discriminative rhetoric" OR "discriminative utterance" OR "discriminative utterances" OR "derogatory speech" OR "derogatory rhetoric" OR "derogatory utterance" OR "derogatory utterances" OR "extreme speech" OR "extreme rhetoric" OR "extreme utterance" OR "extreme utterances" OR "prejudice speech" OR "prejudice rhetoric" OR "prejudice utterance" OR "prejudice utterances" OR "prejudicial speech" OR "prejudicial rhetoric" OR "prejudicial utterance" OR "prejudicial utterances" OR "intolerant speech" OR "intolerant rhetoric" OR "intolerant utterance" OR "intolerant utterances" OR "excitable speech" OR "excitable rhetoric" OR "excitable utterance" OR "excitable utterances" OR "excitable rhetoric" OR "excitable utterance" OR "extremist speech" OR "extremist rhetoric" OR "extremist utterance" OR "extremist utterances" OR "online hate" OR "rhetoric of hate" OR "rhetoric of discrimination" OR "harm speech")) OR AB ((cyberhate OR cyberbully* OR "hate speech" OR "hate rhetoric" OR "hate utterance" OR "hate utterances" OR "hateful speech" OR "hateful rhetoric" OR "hateful utterance" OR "hateful utterances" OR "degrading speech" OR "degrading rhetoric" OR "degrading utterance" OR "degrading utterances" OR "threatening speech" OR "threatening rhetoric" OR "threatening utterance" OR "threatening utterances" OR "harassing speech" OR "harassing rhetoric" OR "harassing utterance" OR "harassing utterances" OR "stigmatizing speech" OR "stigmatizing	Search modes - Boolean/Phrase	954

	<p>rhetoric" OR "stigmatizing utterance" OR "stigmatizing utterances" OR "vile speech" OR "vile rhetoric" OR "vile utterance" OR "vile utterances" OR "offensive speech" OR "offensive rhetoric" OR "offensive utterance" OR "offensive utterances" OR "outrageous speech" OR "outrageous rhetoric" OR "outrageous utterance" OR "outrageous utterances" OR "discriminatory speech" OR "discriminatory rhetoric" OR "discriminatory utterance" OR "discriminatory utterances" OR "discriminating speech" OR "discriminating rhetoric" OR "discriminating utterance" OR "discriminating utterances" OR "harmful speech" OR "harmful rhetoric" OR "harmful utterance" OR "harmful utterances" OR "racist speech" OR "racist rhetoric" OR "racist utterance" OR "racist utterances" OR "discriminative speech" OR "discriminative rhetoric" OR "discriminative utterance" OR "discriminative utterances" OR "derogatory speech" OR "derogatory rhetoric" OR "derogatory utterance" OR "derogatory utterances" OR "extreme speech" OR "extreme rhetoric" OR "extreme utterance" OR "extreme utterances" OR "prejudice speech" OR "prejudice rhetoric" OR "prejudice utterance" OR "prejudice utterances" OR "prejudicial speech" OR "prejudicial rhetoric" OR "prejudicial utterance" OR "prejudicial utterances" OR "intolerant speech" OR "intolerant rhetoric" OR "intolerant utterance" OR "intolerant utterances" OR "excitable speech" OR "excitable rhetoric" OR "excitable utterance" OR "excitable utterances" OR "excitable rhetoric" OR "excitable utterance" OR "extremist speech" OR "extremist rhetoric" OR "extremist utterance" OR "extremist utterances" OR "online hate" OR "rhetoric of hate" OR "rhetoric of discrimination" OR "harm speech")) OR KW ((cyberhate OR cyberbully* OR "hate speech" OR "hate rhetoric" OR "hate utterance" OR "hate utterances" OR "hateful speech" OR "hateful rhetoric" OR "hateful utterance" OR "hateful utterances" OR "degrading speech" OR "degrading rhetoric" OR "degrading utterance" OR "degrading utterances" OR "threatening speech" OR "threatening rhetoric" OR "threatening utterance" OR "threatening utterances" OR "harassing speech" OR "harassing rhetoric" OR "harassing utterance" OR "harassing utterances" OR "stigmatizing speech" OR "stigmatizing rhetoric" OR "stigmatizing utterance" OR "stigmatizing utterances" OR "vile speech" OR "vile rhetoric" OR "vile utterance" OR "vile utterances" OR "offensive speech" OR "offensive rhetoric" OR "offensive utterance" OR "offensive utterances" OR "outrageous speech" OR "outrageous rhetoric" OR "outrageous utterance" OR "outrageous utterances" OR "discriminatory speech" OR "discriminatory rhetoric" OR "discriminatory utterance" OR "discriminatory utterances" OR "discriminating speech" OR "discriminating rhetoric" OR "discriminating utterance" OR "discriminating utterances" OR "harmful speech" OR "harmful rhetoric" OR "harmful utterance" OR "harmful utterances" OR "racist speech" OR "racist rhetoric" OR "racist utterance" OR "racist utterances" OR "discriminative speech" OR "discriminative rhetoric" OR "discriminative utterance" OR "discriminative utterances" OR "derogatory speech" OR "derogatory rhetoric" OR "derogatory utterance" OR "derogatory utterances" OR "extreme speech" OR "extreme rhetoric" OR "extreme utterance" OR "extreme utterances" OR "prejudice speech" OR "prejudice rhetoric" OR "prejudice utterance" OR "prejudice utterances" OR "prejudicial speech" OR "prejudicial rhetoric" OR "prejudicial utterance" OR "prejudicial utterances" OR "intolerant speech" OR "intolerant rhetoric" OR "intolerant utterance" OR "intolerant utterances" OR "excitable speech" OR "excitable rhetoric" OR "excitable utterance" OR "excitable utterances" OR "excitable rhetoric" OR "excitable utterance" OR "extremist speech" OR "extremist rhetoric" OR "extremist utterance" OR "extremist</p>		
--	--	--	--

	utterances" OR "online hate" OR "rhetoric of hate" OR "rhetoric of discrimination" OR "harm speech"))		
S5	S3 OR S4	Search modes - Boolean/Phrase	1,345
S6	DE "MUSLIMS"	Search modes - Boolean/Phrase	4,337
S7	DE "MUSLIMS in non-Islamic countries"	Search modes - Boolean/Phrase	128
S8	DE "ISLAMOPHOBIA"	Search modes - Boolean/Phrase	717
S9	DE "ISLAMIZATION"	Search modes - Boolean/Phrase	96
S10	S6 OR S7 OR S8 OR S9	Search modes - Boolean/Phrase	4,996
S11	TI (muslim* OR islam* OR moslem*) OR AB (muslim* OR islam* OR moslem*) OR KW (muslim* OR islam* OR moslem*)	Search modes - Boolean/Phrase	25,006
S12	S10 OR S11	Search modes - Boolean/Phrase	25,201
S13	DE "ONLINE chat"	Search modes - Boolean/Phrase	176
S14	DE "INTERNET forums"	Search modes - Boolean/Phrase	100
S15	DE "VIRTUAL communities"	Search modes - Boolean/Phrase	506
S16	DE "ONLINE comments"	Search modes - Boolean/Phrase	72
S17	DE "MICROBLOGS"	Search modes - Boolean/Phrase	129
S18	DE "BLOGS"	Search modes - Boolean/Phrase	612
S19	DE "SOCIAL media"	Search modes - Boolean/Phrase	3,698
S20	DE "BACKCHANNELS (Social media)"	Search modes - Boolean/Phrase	7
S21	DE "INSTANT messaging"	Search modes - Boolean/Phrase	103
S22	DE "INTERNET"	Search modes - Boolean/Phrase	7,418
S23	S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22	Search modes - Boolean/Phrase	11,941
S24	TI ((online OR "social media*" OR digital* OR internet* OR Facebook OR blog* OR Twitter OR chat* OR youtube OR WhatsAp* OR Messenger OR "web sites" OR websites OR "social netw*" OR	Search modes - Boolean/Phrase	61,634

	cyber OR cyberspace) OR AB ((online OR "social media*" OR digital* OR internet* OR Facebook OR blog* OR Twitter OR chat* OR youtube OR WhatsAp* OR Messenger OR "web sites" OR websites OR "social netw*" OR cyber OR cyberspace) OR KW ((online OR "social media*" OR digital* OR internet* OR Facebook OR blog* OR Twitter OR chat* OR youtube OR WhatsAp* OR Messenger OR "web sites" OR websites OR "social netw*" OR cyber OR cyberspace)		
S25	S23 OR S24	Search modes - Boolean/Phrase	62,784
S26	S5 AND S12 AND S25	Search modes - Boolean/Phrase	19
S27	S5 AND S12 AND S25	Limiters - Date of Publication: 20100101-20201231 Search modes - Boolean/Phrase	17
S28	S5 AND S12 AND S25	Limiters - Date of Publication: 20100101-20201231 Search modes - Boolean/Phrase	17

Database: Scopus
Dato: 09.11 2020
Treff: 57

(TITLE-ABS-KEY (muslim* OR islam* OR moslem*)) AND (TITLE-ABS-KEY (cyberhate OR cyberbully* OR "hate speech" OR "hate rhetoric" OR "hate utterance" OR "hate utterances" OR "hateful speech" OR "hateful rhetoric" OR "hateful utterance" OR "hateful utterances" OR "degrading speech" OR "degrading rhetoric" OR "degrading utterance" OR "degrading utterances" OR "threatening speech" OR "threatening rhetoric" OR "threatening utterance" OR "threatening utterances" OR "harassing speech" OR "harassing rhetoric" OR "harassing utterance" OR "harassing utterances" OR "stigmatizing speech" OR "stigmatizing rhetoric" OR "stigmatizing utterance" OR "stigmatizing utterances" OR "vile speech" OR "vile rhetoric" OR "vile utterance" OR "vile utterances" OR "offensive speech" OR "offensive rhetoric" OR "offensive utterance" OR "offensive utterances" OR "outrageous speech" OR "outrageous rhetoric" OR "outrageous utterance" OR "outrageous utterances" OR "discriminatory speech" OR "discriminatory rhetoric" OR "discriminatory utterance" OR "discriminatory utterances" OR "discriminating speech" OR "discriminating rhetoric" OR "discriminating utterance" OR "discriminating utterances" OR "harmful speech" OR "harmful rhetoric" OR "harmful utterance" OR "harmful utterances" OR "racist speech" OR "racist rhetoric" OR "racist utterance" OR "racist utterances" OR "discriminative speech" OR "discriminative rhetoric" OR "discriminative utterance" OR "discriminative utterances" OR "derogatory speech" OR "derogatory rhetoric" OR "derogatory utterance" OR "derogatory utterances" OR "extreme speech" OR "extreme rhetoric" OR "extreme utterance" OR "extreme utterances" OR "prejudice speech" OR "prejudice rhetoric" OR "prejudice utterance" OR "prejudice utterances" OR "prejudicial speech" OR "prejudicial rhetoric" OR "prejudicial utterance" OR "prejudicial utterances" OR "intolerant speech" OR "intolerant rhetoric" OR "intolerant utterance" OR "intolerant utterances" OR "excitable speech" OR "excitable rhetoric" OR "excitable utterance" OR "excitable utterances" OR "excitable rhetoric" OR "extremist speech" OR "extremist rhetoric" OR "extremist utterance" OR "extremist utterances" OR "online hate" OR "rhetoric of hate" OR "rhetoric of discrimination" OR "harm speech")) AND (TITLE-ABS-KEY (online OR "social media*" OR digital* OR internet* OR facebook OR blog* OR twitter OR chat* OR youtube OR whatsapp* OR messenger OR "web sites" OR websites OR "social netw*" OR cyber OR cyberspace) AND (LIMIT-TO (SRCTYPE , "j") OR LIMIT-TO (SRCTYPE , "p")) AND (LIMIT-TO (LANGUAGE , "English")) View less

-



Database: Open Grey

Dato: 11.12 2020

Antall treff: 47 – ikke eksportert til Endnote pga grensesnitt.

("hate speech" OR "hate rhetoric" OR "hate utterance" OR "hate utterances" OR "hateful speech" OR "hateful rhetoric" OR "hateful utterance" OR "hateful utterances" OR "degrading speech" OR "degrading rhetoric" OR "degrading utterance" OR "degrading utterances" OR "threatening speech" OR "threatening rhetoric" OR "threatening utterance" OR "threatening utterances" OR "harassing speech" OR "harassing rhetoric" OR "harassing utterance" OR "harassing utterances" OR "stigmatizing speech" OR "stigmatizing rhetoric" OR "stigmatizing utterance" OR "stigmatizing utterances" OR "vile speech" OR "vile rhetoric" OR "vile utterance" OR "vile utterances" OR "offensive speech" OR "offensive rhetoric" OR "offensive utterance" OR "offensive utterances" OR "outrageous speech" OR "outrageous rhetoric" OR "outrageous utterance" OR "outrageous utterances" OR "discriminatory speech" OR "discriminatory rhetoric" OR "discriminatory utterance" OR "discriminatory utterances" OR "discriminating speech" OR "discriminating rhetoric" OR "discriminating utterance" OR "discriminating utterances" OR "harmful speech" OR "harmful rhetoric" OR "harmful utterance" OR "harmful utterances" OR "racist speech" OR "racist rhetoric" OR "racist utterance" OR "racist utterances" OR "discriminative speech" OR "discriminative rhetoric" OR "discriminative utterance" OR "discriminative utterances" OR "derogatory speech" OR "derogatory rhetoric" OR "derogatory utterance" OR "derogatory utterances" OR "extreme speech" OR "extreme rhetoric" OR "extreme utterance" OR "extreme utterances" OR "prejudice speech" OR "prejudice rhetoric" OR "prejudice utterance" OR "prejudice utterances" OR "prejudicial speech" OR "prejudicial rhetoric" OR "prejudicial utterance" OR "prejudicial utterances" OR "intolerant speech" OR "intolerant rhetoric" OR "intolerant utterance" OR "intolerant utterances" OR "excitable speech" OR "excitable rhetoric" OR "excitable utterance" OR "excitable utterances" OR "extremist speech" OR "extremist rhetoric" OR "extremist utterance" OR "extremist utterances" OR "online hate" OR "rhetoric of hate" OR "rhetoric of discrimination" OR "harm speech")

Language: English

Database: Swepub.kb.se

Dato: 10.12 2020

Antall treff: 35

(näthat OR näthot OR näthets OR "hat online" OR "hat på nätet" OR "nätkränkningar" OR ((hatfylld* OR hatfull* OR rasistisk* OR främlingsfientlig*) AND (kommentar* OR yttring* OR kränkning* OR budskap)) OR "hatiskt tal" OR "hatpropaganda" OR "hatets språk" OR hatretorik OR "hets mot folkgrupp" OR "hate speech" OR hetspropaganda OR ärekränk*) conttype:ref

Alt innhold, ikke avgrenset på år. «Swepub inneholder referenser til de forskningspublikasjoner som finns registrerade i för närvarande ett fyrtiotal svenska lärosätens och andra myndigheters publiceringsdatabaser. Urval och omfattning varierar bland deltagande lärosäten och myndigheter.

Både publicerade och opublicerade arbeten tas med och samma publikation kan förekomma i olika versioner till exempel både som konferensbidrag och som tidskriftsartikel.» swepub.kb.se/help.jsp

Database: Diva-portal.org
Dato: 10.12.20 20
Antall treff: 95

(näthat OR näthot OR näthets OR "hat online" OR "hat på nätet" OR "nätkränkningar" OR "hatfylld*" OR "hatiskt tal" OR "hatpropaganda" OR "hatets språk" OR hatretorik OR "hets mot folkgrupp" OR "främlingsfientlig*" OR "hate speech" OR "rasistisk*" OR hetspropaganda OR ärekränk*)

Publiseringsår: 2010-2020

Database: Den danske forskningsdatabase (DEF)
Dato: 10.12 2020
Treff: 20

(hate speech) OR (nedladende ytring*) OR (racistiske ytring*) OR (fremmedfjendsk* ytring*) OR (hadefuld* ytring*) OR (hadetale*) OR (hadefuld* besked*) OR (hadske kommentar*) OR (hadbesked*) OR (hadefuld* kommentar*) OR (internethad) OR (hadefuld? sprog*) OR (hadefuld retorik) OR (nethad) OR (hadefuld* tale*) OR (hadesytring*) OR (krænkende ytring*) OR (internettrus*) OR (net-trus*) OR (ærekrænkende ytring*) OR (ærekrænkende kommentar*) OR (hetz mod folkegruppe*) OR (racistisk* kommentar*) OR (diskriminerende ytring*)
Udgivelsesår:2010 to 2020, publication type: article

Database: Oria

Dato: 11.12 2020

Antall treff: 14

Alle felt inneholder (hatefulle ytring?) OR (hatefull ytring) hatretorikk OR hatprat OR netthat OR netthets OR netthetsing

AND

Alle felt inneholder (muslim* OR islam*)

År: 2010-2020

Database: Google Scholar (scholar.google.no)

Dato: 11.12 2020

"hate AROUND(3) islam" , 2010 – 2020: **718 TREFF**

"Hatefulle" AROUND(3) islam", 2010 – 2020: **228 treff**

Database: Google Scholar (scholar.google.dk)

Dato: 11.12 2020

hadefuld AROUND(3) islam, 2010-2020: **antall treff 187**

Database: Google Scholar (scholar.google.se)

Dato: 11.12 2020

näthat AROUND(3) islam: **antall treff 79**