FORSVARET

Forsvarets høgskole

Etterretningsanalyse og stordata

Stordatadrevet ACH – bedre analyser eller enda en tidstyv?

Lars Roar Uggerud Dugstad

Masteroppgave

Forsvarets høgskole

vår 2021

Forord

Denne masteroppgaven er blitt til som en del av masterstudiet ved Forsvarets Høgskole 2019-2021. Arbeidet med oppgaven har vært utrolig givende, og har gitt muligheter for fordypning i et komplekst tema som har gitt ny kunnskap og innsikt. Det er en lang og til tider krevende reise, som nå endelig går mot slutten.

Denne oppgaven hadde ikke vært mulig å gjennomføre uten svært verdifulle bidrag underveis. For at oppgaven fremstår som den gjør rettes en stor takk til begge mine veiledere. Hovedveileder Mass Soldal Lund har gitt svært verdifulle bidrag underveis, og sørget for at prosjektet har holdt stødig kurs hele veien. Hans metodekunnskap og evne til å tenke helhetlig har løftet oppgaven. Biveileder, sjefsforsker ved FFI, Bjørn Jervell Hansen har også hatt en nøkkelrolle. Hans støtte, tålmodighet og veiledning i oppsett og drift av stordatainfrastrukturen har vært uvurderlig for at denne oppgaven har vært mulig. Uten deg hadde dette blitt en teoretisk studie, det hadde ikke gitt samme læringsutbytte, eller vært like gøy! Tusen takk rettes også til Merete Ruud som har lest gjennom oppgaven og gitt verdifulle tilbakemeldinger på det språklige.

Oppgaven står for egen regning. Analysen og vurderingene er mine egne.

Sist, men ikke minst vil jeg rette en takk til min kone Kari for hennes tålmodighet, og for at hun har gitt meg mulighet til å lukke meg inne på kontoret all ledig fritid de siste månedene før innlevering. Takk for at dette var mulig!

Sammendrag

Ved å ta i bruk nye teknologier tilknyttet stordataanalyse, hevder denne oppgaven at det finnes et stort potensial for å kunne gjøre etterretningsanalyse raskere, og sette den enkelte analytiker i stand til å dekke over en større informasjonsmengde med høyere presisjon. Dette muliggjøres av den teknologiske utviklingen innenfor databehandling og stordatasystemer som gjør det mulig å overføre, analysere og sammenstille informasjon raskere og mer effektivt, og gjennom dette kunne tolke store datamengder.

Utgangspunktet for denne oppgaven var en observasjon av at stordata- og etterretningsanalyse har en del fellestrekk, og at stordataanalyse derfor kan ha et potensiale for å bidra til mer effektiv etterretningsanalyse. Målsetningen ble derfor å utforske om, og på hvilken måte, stordataanalyse kan understøtte etterretningsanalyse. For å gjøre dette blir det foreslått en metode som kombinerer stordata- og etterretningsanalyse. Metoden har fått navnet *stordatadrevet ACH*. Metoden benytter stordataanalyse til å avdekke mønstre og foreslå konklusjoner, mens ACH benyttes som et rammeverk for å utvikle hypoteser, vurdere kontekst og ta de endelige beslutningene. Metoden blir evaluert gjennom et eksperiment.

Eksperimentet tester en hypotese om at det foregår ulovlig, urapportert eller uregulert (UUU) fiske i norske farvann. Med bakgrunn i tre indikatorer for UUU-fiske ble det beskrevet algoritmer som kunne svare på de nevnte indikatorene. For å gjennomføre eksperimentet ble det etablert en stordatainfrastruktur, og Kystverkets åpne AIS-strøm ble benyttet som stordatakilde. I løpet av eksperimentet ble mer enn 5 000 000 AIS-meldinger analysert, og svar på de ulike indikatorene ble presentert i sanntid. Resultatet av eksperimentet var at det ble identifisert *ett* fiskefartøy som vi kan hevde at det er økt sannsynlighet for at driver med UUU-fiske.

Et sentralt funn er at det er tidkrevende og komplisert å etablere stordataløsninger. Stordatainfrastruktur krever spesifikk kompetanse både for å sette opp, konfigurere og drifte. Det er derfor viktig at det etableres et tett samarbeid mellom etterretningsanalytiker og dataingeniør(er) når det skal etableres slike løsninger.

Studien viser at *stordatadrevet ACH* er et kraftig verktøy når det anvendes riktig. Den endelige konklusjonen er derfor at *stordatadrevet ACH* kan øke omfanget og anvendeligheten til ACH spesielt og til etterretningsanalyse generelt, men dette forutsetter at det benyttes på rett problemstillinger. Det må ses på som et supplement til eksisterende prosesser og systemer, ikke en erstatning.

Summary

By using new technologies, and especially big data analysis, this thesis claims that there is a great potential to do intelligence analysis faster and enable the individual analyst to cover a larger amount of information with higher precision. This is made possible by the technological development within data processing and big data systems that make it possible to transfer, analyse and compile information faster and more efficiently, and thus be able to interpret large amounts of data.

The starting point for this thesis was an observation that big data and intelligence analysis have a number of common features, and that big data analysis therefore may have a potential to contribute to more efficient intelligence analyses. The aim was therefore to explore whether and in what way big data analysis could support intelligence analysis. To do this, a method is proposed that combines big data and intelligence analysis. The method has been named *big data driven ACH*. The method uses big data analysis to uncover patterns and propose conclusions, while ACH is used as a framework for developing hypotheses, assessing context and making the final decisions. The method is evaluated through an experiment.

The experiment tests a hypothesis that there is illegal, unreported or unregulated (IUU) fishing in Norwegian territorial waters. Based on three indicators for IUU fishing, algorithms were developed that could collect the mentioned indicators. To carry out the experiment, a big data infrastructure was established and the Norwegian Coastal Administration's open AIS stream was used as a big data source. During the experiment, more than 5,000,000 AIS messages were analysed, and answers to the various indicators were presented in real time. The result of the experiment was that *one* fishing vessel was identified which we may claim is more likely to engage in IUU fishing.

A key finding is that it is time-consuming and complicated to establish big data solutions. Big data infrastructure requires specific expertise to set up, configure and operate. It is therefore important that a close collaboration is established between the intelligence analyst and the computer engineer(s) when establishing such solutions.

The study shows that *big data driven ACH* is a powerful tool when used correctly. The final conclusion is therefore that *big data driven ACH* can increase the scope and applicability of ACH in particular and for intelligence analysis in general, but this assumes that it is used on the right tasks. Lastly, it must be seen as a supplement to existing processes and systems, not a replacement.

Innholdsfortegnelse

1 Innledning

Den teknologiske utviklingen innenfor databehandling og stordatasystemer skjer nå i så høyt tempo og er av et slikt omfang at flere hevder vi står inne i en ny teknologisk revolusjon (Forsvarsdepartementet, 2019; Froelich, 2020). Direktøren ved Forsvarets Forskningsinstitutt (FFI) uttaler følgende:

> Fremskritt innen områder som datakraft, kunstig intelligens og stordata legger grunnlag for fantastiske fremskritt i alt fra helse til finans, industri og transport, men få områder vil påvirkes sterkere enn forsvarssektoren. Her skjer endringene raskt, med mer dyptgripende og vidtrekkende konsekvenser enn vi har vært vant til (Størdal, 2019).

Dette utsagnet underbygges også av flere militære trendstudier som er samstemte om hvilke teknologiområder som vil ha stor og økende betydning for militære operasjoner framover (Beadle, Diesen, Nyhamar & Bostad, 2019, s. 46; Skjelland et al., 2019, s. 22). Her fremheves evnen til å tolke enorme datamengder for å etablere et overlegent situasjonsbilde. Den teknologiske utviklingen innen informasjons- og kommunikasjonsteknologi vil gjøre det mulig å overføre, analysere og sammenstille informasjon raskere og mer effektivt. Automatiserte analyser gjør det i prinsippet mulig å håndtere økte informasjonsmengder av ulike typer og fra ulike kilder, noe som kan bidra til deteksjon av objekter og hendelser og gi mer nøyaktige måldata (Forsvarets forskningsinstitutt, 2019, s. 10).

Ny teknologi, særlig smarttelefonen, gjør også at informasjon spres stadig raskere og formidles hurtigere enn noen gang tidligere (Stenslie, Haugom & Vaage, 2019, s. 24-25). *Information overload* er således beskrivende for den tiden vi lever i. Begrepet brukes for å beskrive hvor vanskelig det er å forstå et problem og effektivt ta beslutninger når en har for mye informasjon. Slike situasjoner oppstår når mengden av informasjon overgår evnen til å prosessere denne (Speier, Valacich & Vessey, 1999). For å kunne nyttiggjøre oss av all denne informasjonen vi har tilgjengelig må den derfor tolkes eller analyseres og settes i en eller annen kontekst (Vivento & Kaupang, 2015, s. 15). Først da kan informasjon bli til kunnskap, og dermed anvendbare i sivile eller militære beslutningsprosesser.

Å tolke, analysere og sette data i rett kontekst er en sentral del av arbeidet til etterretningsanalytikeren. Utfordringene med den økende informasjonsmengden og de utfordringer som ligger i å få prosessert disse, har lenge være erkjent i etterretningskretser. Allerede i 2013 blir det gjort et poeng av dette i den norske Etterretningsdoktrinen: «*I dagens informasjonssamfunn er*

det sjelden knapphet på informasjon som er utfordringen, men heller at informasjonstilfanget er så stort at utfordringen blir å velge hva som skal vektlegges» (Etterretningstjenesten, 2013).

Utgangspunktet for denne oppgaven var en observasjon av at stordata- og etterretningsanalyse har en del fellestrekk, og at stordataanalyse derfor kan ha et potensiale for å bidra til mer effektiv etterretningsanalyse. Målsetningen ble derfor å utforske om, og på hvilken måte, stordataanalyse kan understøtte etterretningsanalyse. Ved å ta i bruk nye teknologier tilknyttet stordataanalyse, hevder denne oppgaven at det er et stort potensial i å kunne gjøre etterretningsanalyse raskere og muliggjøre at den enkelte analytiker kan behandle en større informasjonsmengde med høyere presisjon. ß

Datamaskiner er overlegne oss mennesker når det gjelder å behandle store informasjonsmengder raskt, og med et pålitelig resultat. Vi mennesker mister fort tålmodigheten, kan bli påvirket av biaser[1] eller rett og slett overse viktig informasjonsbiter, noe som gjør at det endelige resultatet kan være mindre pålitelig. Samtidig har vi vår styrke i at vi har fantasi og kan forestille oss ting som ikke har skjedd. Dette gjør at vi kan utvikle forskjellige scenarier og hypoteser, og finne nye løsninger på ukjente problem (Haugom, Hemmingby & Pedersen, 2019, s. 102). Det ligger derfor et potensiale i å utnytte de sterke sidene til både menneske og maskin. Maskinen kan settes til å avdekke mønstre og foreslå konklusjoner, med andre ord foredle rådata til nyttig informasjon (Stolpe, Hansen & Halvorsen, 2019, s. 12-13). Maskinen vil hjelpe analytikeren med å håndtere de enorme informasjonsmengdene, trekke ut det som er relevant, og å få presentert denne informasjonen i strukturert form. Analytikeren vil utvikle hypoteser, se helheten, vurdere kontekst og ta de endelige beslutningene.

En metode som derfor er interessant å bruke som utgangspunkt for å kombinere etterretnings- og stordataanalyse, er *«Analysis of Competing Hypotheses»* (ACH). Metoden skal hjelpe analytikeren med å identifisere et sett med ulike, alternative hypoteser og gjennomføre en systematisk evaluering av disse oppimot tilgjengelige bevis. Resultatet av prosessen er en vurdering av hvilken hypotese som er mest sannsynlig med utgangspunkt i hvilken hypotese som i minst grad er avkreftet (Artner, Girven & Bruce, 2016, s. 2; Heuer & Pherson, 2010, s. 160). Metoden egner seg for de fleste problemstillinger hvor det er flere forskjellige mulige forklaringsalternativer. ACH er spesielt god når det er en stor datamengde man kan analysere og evaluere (Heuer & Pherson, 2010, s. 160).

[1] Bias innebærer at resultater eller slutninger er skjeve eller feilaktige, ved at de avviker systematisk fra de virkelige forholdene som utforskes (Grønmo, 2020).

Richards Heuer utviklet metoden mens han jobbet i CIA på 80-tallet. Her observerte han at analytikere ofte kun utarbeidet en hypotese og deretter søkte etter bevis som støttet denne. Dette er en behagelig og tidseffektiv arbeidsmetodikk da det sparer mye tid og arbeid, og veldig ofte er det en trygg tilnærming da fremtiden ofte er en forlengelse av fortiden. Allikevel er dette tankesettet svært problematisk (Heuer, 2005, s. 3). I en militær setting hvor man forholder seg til en motstander som til enhver tid vil forsøke å tilrive seg overtaket, må man forberede seg på det uventede og ta grep for å unngå overraskelser.

En av hovedutfordringene ved teknikken er at det er arbeidskrevende å utarbeide mange hypoteser, for deretter å jobbe systematisk med å bekrefte eller avkrefte disse (Coulthart, 2016; Marrin, 2002; Stenslie, 2019). På grunn av dagens tilgjengelige verktøy og arbeidsmetoder, som baserer seg på Word, Excel, Powerpoint og tilsvarende program, er arbeidet med ACH manuelt og krever betydelig arbeidsinnsats fra den enkelte analytiker. Dette gjør at arbeidet med slike teknikker paradoksalt nok nedprioriteres i hektiske perioder (Stenslie, 2019, s. 77), selv om det er i nettopp slike perioder at teknikken har sin styrke med en tydelig vekting av bevis og gjennom dette reduksjon av biaser og magefølelse, og som et resultat av dette mulighet for en mer objektiv analyse. Denne oppgaven presenterer derfor en metode som kombinerer ACH og stordataanalyse, og som undersøker om denne kombinasjonen vil gjøre ACH mer anvendelig for analytikeren. Ved å kombinere datamaskinens presisjon og menneskets kreativitet er ambisjonen å kunne analysere større datamengder med høyere presisjon.

1.1 Problemstilling og avgrensning

Målsetningen med dette arbeidet er å bidra til økt kunnskap om hvordan bruken av stordataanalyse kan understøtte etterretningsanalyse, og hvordan disse kan kombineres for å gjøre etterretningsanalyse mer effektiv.

Problemstillingen for arbeidet er:

«Vil integrering av sanntids stordataanalyse i «Analysis of competing hypotheses» øke omfanget og anvendeligheten av analysemetoden?»

Problemstillingen vil bli analysert ved å svare på følgende forskningsspørsmål:

Forskningsspørsmål 1: Hvordan kan stordataanalyse og ACH integreres metodisk?

Forskningsspørsmål 2: Hvordan kan stordataanalyse og ACH integreres teknisk?

Forskningsspørsmål 3: Tilfører metoden merverdi og vil den gjøre at analytikeren kan bruke mer tid på analyse?

For å få svar på dette presenterer studien et forslag til en ny metode for å kombinere disse kalt *stordatadrevet ACH og* et eksperiment med bruken av denne metoden.

For å utnytte de mulighetene som teknologien gir var det ønskelig å se på muligheten for å presentere svar på analyser i nær sanntid. Dette medførte at oppgaven avgrenses til å se på såkalte strømmesystemer, da disse er spesielt egnet til slike arbeidsoppgaver. Videre var det et ønske om å holde oppgaven ugradert. Dette gjorde det derfor mindre aktuelt å bruke militære problemstillinger og datakilder fra Forsvarets systemer. Av den grunn ble det valgt å benytte en problemstilling i eksperimentet knyttet til såkalt ulovlig, uregulert og urapportert (UUU) fiske, og å benytte Automatic Identification System (AIS) som stordatakilde.

1.2 Oppgavens struktur

Oppgaven er todelt. Første del består av kapittel to til fem og utgjør oppgavens teoretiske rammeverk. Andre del består av kapittel seks til elleve, og inneholder en beskrivelse av oppgavens eksperiment, en analyse og drøfting av resultater og konklusjon.

Kapittel to vil redegjøre for oppgavens valg av forskningsdesign og metodisk fremgangsmåte. Her presenteres også målekriteriene som *stordatadrevet ACH* og tilhørende stordatainfrastruktur skal måles mot.

Kapittel tre og fire er teoretiske gjennomganger av hhv. etterretningsanalyse og stordataanalyse. Det fokuseres på sentrale begrep og konsepter, slik at det dannes et grunnlag for å forstå metoden som blir presentert i kapittel fem.

Kapittel fem presenterer nyskapning i denne oppgaven, *stordatadrevet ACH*. Her redegjøres det for hvordan stordata- og etterretningsanalyse kombineres ved å presentere en metodebeskrivelse.

Kapittel seks presenterer den valgte stordatainfrastrukturen som brukes til å teste metoden som ble presentert i kapittel fem.

Kapittel syv beskriver grunnleggende aspekter ved konseptet og datagrunnlaget for stordatakilden som benyttes i eksperimentet, Automatic Identification System (AIS). Denne delen viser også til mulige feilkilder.

Kapittel åtte redegjør for selve eksperimentet og gjennomføringen av det. Kapittelet viser hvordan *stordatadrevet ACH* brukes for å bryte ned et problemsett knyttet UUU-fiske. Det presenteres en stegvis gjennomgang av metoden fra problemstilling, via indikatornedbryting, til innhenting og presentasjon.

Kapittel ni presenterer resultatene fra eksperimentet samt hvilke svar stordatainfrastrukturen ga oss på problemsettet knyttet til UUU-fiske.

Kapittel ti utgjør oppgavens diskusjonsdel. Her vil forskningsspørsmålene drøftes, med fokus på om metoden har truffet målekriteriene beskrevet i metodekapittelet.

Kapittel elleve presenterer oppgavens konklusjon. Her oppsummeres oppgaven og de viktigste funnene. Det anbefales også noen temaer som kan være interessante for videre forskning med utgangspunkt i denne oppgaven.

2 Metode

Vitenskapsteorien er opptatt av at forskeren skal prøve å synliggjøre sitt metodiske og empiriske ståsted for seg selv og leseren. Forskeren skal vise med hvilke begreper og på hvilken måte hen ser og beskriver verden. Forskningen skal være skapt på en systematisk, kritisk, nøyaktig og troverdig måte, og etter bestemte metoder. Videre skal vitenskapelig kunnskap være basert på systematisk kritisk refleksjon. Dermed er det innhenting og håndtering av kunnskapen som bestemmer hvorvidt den holder mål i forhold til en relevant praksis, og derav blir metoden helt essensiell (Bergander & Johnsen, 2006, s. 18).

For å svare på problemstillingen i denne oppgaven vil jeg benytte meg av teknologivitenskapelig metode. Stølen (2019, s. 15) definerer teknologivitenskap som *«vitenskap der man fokuserer på å utvide virkeligheten med nye eller vesentlige bedre artefakter.»* Fokuset er altså på artefakter som er menneskeskapte ting, fenomener og objekter. Dette begrepet dekker alt fra kniver til satellitter i bane (Stølen, 2019, s. 9-11). Begrepet teknologi har også mange definisjoner, men denne oppgaven legger følgende definisjon til grunn: *«læren om og studiet av praktiske framgangsmåter i håndverk og industri»* (Bokmålsordboka, 2021).

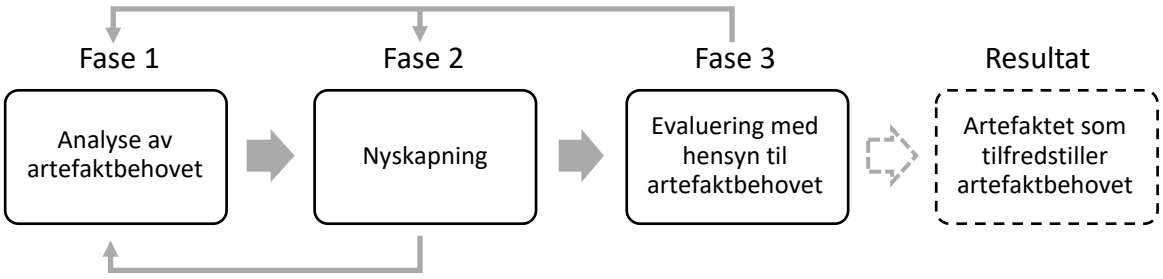
Stølen (2019) hevder at når man jobber med utvikling av ny teknologi må man tilnærme seg dette med «*andre metodiske briller»* enn i de klassiske vitenskapene. Istedenfor å dele verden i de to vanlige kategoriene samfunnsvitenskap og naturvitenskap, velger Stølen å dele verden i teknologivitenskap og forklaringsvitenskap. Teknologivitenskap er vitenskapen om å skape, oppfinne eller frembringe nye ting i form av menneskeskapte objekter, såkalte artefakter, ofte omtalt som teknologi. Innen teknologivitenskap er forskningsresultatet alltid et nytt eller forbedret artefakt. Stølen bruker begrepet forklaringsvitenskap om naturvitenskap, samfunnsvitenskap og annen vitenskap som handler om å forstå verden som den er, inkludert artefaktene som allerede eksisterer. Fokuset i forklaringsvitenskapen er å beskrive, forstå og karakterisere lovmessigheter i den virkeligheten som omgir oss, ikke skape ny teknologi.

Denne oppgaven presenter et forslag til en metode for å kombinere etterretningsanalyse og stordata. Dette er utnyttelse av teknologi og er en menneskeskapt ting eller idé. Sånn sett skapes det et artefakt etter Stølens beskrivelse, og vi velger å benytte hans metodiske rammeverk. Hele grunnlaget bak hans metodebeskrivelse ligger i at klassiske samfunnsvitenskapelige og naturvitenskapelige metoder ikke er like godt egnet til å utvikle ny teknologi (Stølen, 2019, s. 1-2).

2.1 Teknologivitenskapens fremgangsmåte

Som tidligere nevnt, innen teknologivitenskapen er forskningsresultatet alltid et nytt eller forbedret artefakt. Dette kan være seg et nytt robotdesign, nytt dataprogram eller et nytt opplegg for pasientbehandling. Som regel er leveransen et design eller en prototype som kan implementeres eller videreutvikles. Det er det i denne oppgaven også ved at det demonstreres en metode man kan benytte for å integrere etterretningsanalyse og stordataanalyse. Dette kan karakteriseres som en prototype, altså at det bygges en modell av artefaktet (Stølen, 2019, s. 54).

Utgangspunktet i teknologivitenskapen er at man har et behov, en idé eller ser en mulighet for bruken av ny teknologi. Dette kaller vi et artefaktbehov. Behovet for et nytt artefakt kan defineres av forskeren selv eller av andre eksterne som bedrifter, enkeltpersoner, organisasjoner eller det offentlige. Når man har definert artefaktbehovet er det opp til forskerne å tilfredsstille dette behovet (Stølen, 2019, s. 19-20). Selve artefaktbehovet vil normalt beskrives med noen suksesskriterier som brukes til å måle om man har lykkes (Stølen, 2019, s. 41-43). Artefaktbehovet i denne oppgaven springer ut fra forfatterens eget hode, og har oppstått med bakgrunn i egen erfaring med etterretningsanalyse, samt nysgjerrighet for ny teknologi og da spesielt stordataanalyse.



Figur 1: Fremgangsmåte for teknologivitenskap

Oppgaven benytter Stølens fremgangsmåte, eller metode, for teknologivitenskap som vist i figuren over. Som figuren viser kan vi dele denne metoden i tre faser som til slutt skal lede frem til et resultat, altså et artefakt som tilfredsstiller artefaktbehovet. Fasene følger logisk etter hverandre, men prosessen er iterativ, så det er naturlig at man hopper frem og tilbake mellom fasene i løpet av forskningsperioden.

I *fase 1* skal artefaktbehovet analyseres. Her beskrives et potensielt behov for et nytt eller forbedret artefakt. I f*ase 2* skal selve nyskapningen skje. Det skapes, oppfinnes eller lages et artefakt som tilfredsstiller artefaktbehovet. Dette er den kreative fasen og her formuleres hypoteser og nye ideer klekkes ut. Noen ganger er nyskapningen beskjeden, men likevel nyttig ved for eksempel adaptering av et eksisterende artefakt for å dekke et nytt behov. Andre ganger kan nyskapningen være

banebrytende. I *fase 3* skal man analysere nyskapningen. Forskeren evaluerer om artefaktet tilfredsstiller det identifiserte behovet. Dette vil ofte innebære flere eksperimenter eller ulike former for undersøkelser. Gitt at evalueringen er vellykket kan man hevde å ha lykkes. Samtidig, hvis resultatene spriker må man gå tilbake til start og på nytt analysere om man har forstått artefaktbehovet eller om man ikke har lykkes med nyskapningen og må utarbeide et helt nytt artefakt (Stølen, 2019, s. 20).

2.2 Metoden i denne oppgaven

Problemanalysen som danner grunnlaget for en analyse av artefaktbehovet (Fase 1) er dokumentert i kapittel 1 og leder fram til problemstillingen for oppgaven: *Vil integrering av sanntid stordataanalyse i «Analysis of competing hypotheses» (ACH) øke omfanget og anvendeligheten av analysemetoden.* Sett opp mot Stølens beskrivelse av et artefaktbehov er ikke dette en god artefaktbeskrivelse. Et artefaktbehov skal formuleres slik at det beskriver hva dette nye artefaktet skal gjøre. Det er videre vanlig å beskrive det som et suksesskriterium. Dette bør formuleres overordnet, men som et antall relativt presise krav. Det vil gjøre det enklere å måle og analysere resultatene i ettertid (Stølen, 2019, s. 43).

Med utgangspunkt i problemstillingen for oppgaven konkretiseres artefaktbehovet til å være følgende: *Et verktøy som i sanntid analyserer en kontinuerlig stordatastrøm, og fortløpende sjekker om disse dataene tilfredsstiller et sett med forhåndsdefinerte indikatorer og hypoteser, og svarer dette til bruker fortløpende.* Med sanntid menes at dataene prosesseres så hurtig som mulig etter at verktøyet har lest og analysert dataene. Med bruker menes den som benytter verktøyet i sitt daglig virke.

Fase 2, selve nyskapningen, er dokumentert i kapittel 5 og 6. Her skapes selve artefaktet som er et verktøy som søker å løse artefaktbehovet ovenfor. Dette verktøyet består av to deler. Den første delen er en revidert metodebeskrivelse for ACH som gjør det metodisk mulig å integrere stordataanalyse med ACH, dette kalles stordatadrevet ACH. Her tar man utgangspunkt i Heuers opprinnelige ACH-metode og integrerer stordataanalyse. Dette er beskrevet i kapittel 5. Den andre delen er å identifisere en programvarepakke som skal gjennomføre stordataanalysedelen i den reviderte metodebeskrivelsen. Det finnes i dag ingen generisk programvarepakke som kan anvendes på et hvilket som helst stordataproblem uten noen form for tilpasning. Det må derfor gjøres en analyse av behov og datasett, og utfra dette velges noen programvarekomponenter. Den valgte programvarepakken vil bestå av en rekke ulike delkomponenter og vil i sum utgjøre en stordatainfrastruktur. Denne er beskrevet nærmere i kapittel 6.

Fase 3, evalueringen av artefaktbehovet, er dokumentert i kapittel 8 og 9. Verktøyet, både den stordatadrevet ACH og den identifiserte datainfrastrukturen ble testet. Evalueringen ble gjennomført som et felteksperiment. I et felteksperiment testes artefaktet ut under normale betingelser i sine tiltenkte omgivelser, men et mindre antall parametere kan manipuleres eller endres (Stølen, 2019, s. 94-95). Et felteksperiment er realistisk, men er ikke særlig presist fordi det er mange lite kontrollerbare faktorer som potensielt påvirker resultatet (Stølen, 2019, s. 53-54).

Eksperimentet ble gjennomført ved å etablere en stordatainfrastruktur og benytte AIS-data som stordatakilde. Videre ble det identifisert en problemstilling knyttet til ulovlig, urapportert og uregulert (UUU) fiske. Denne problemstillingen gjorde det mulig å teste den reviderte ACH-metoden ved hjelp av AIS-data. I tillegg var det tilgjengelig gode indikatorer for UUU-fiske som kunne videreutvikles til algoritmer som datainfrastrukturen kunne løse. Ved å benytte den reviderte ACH-metoden som utgangspunkt, jobbet jeg meg gjennom stegene i metoden og som resultat ble det utarbeidet algoritmer som besvarte problemsstillingen. Dette er dokumentert i kapittel 8. Videre ble løsningen testet på en sanntids AIS-datastrøm i omtrent 36 timer. Underveis i prosessen ble algoritmer og noen andre innstillinger/parametere justert fortløpende, enten for å øke ytelse eller kvalitet på resultatet. Resultatene fra eksperimentet presenteres i kapittel 9. Disse resultatene legger grunnlag for en diskusjon i kapittel 10 der nyskapningen evalueres.

2.3 Etiske vurderinger

Eksperimentet vil identifisere et antall enkelfartøy som har utslag på en eller flere indikatorer for UUU-fiske og derfor ble prosjektet meldt inn til personvernombudet for forskning ved Norsk Senter for Forskningsdata (NSD).

Eksperimentets ambisjon er å teste den reviderte ACH-metodikken, således er ikke fokuset å avsløre fartøy som har økt sannsynlighet for UUU-fiske. Det er ikke eksperimentets hensikt å rette mistanke mot enkeltfartøyer, deres besetninger eller eiere. Problemstillingen knyttet til UUU-fiske brukes først og fremst fordi den egner seg til å teste metoden ved hjelp av de verktøyene og det datasettet som oppgaven benytter. I tillegg er dette tema som det er gjort en del forskning på. Det gjorde at det var mulig å finne gode kilder til utviklingen av algoritmene, slik at det ble gjennomført et eksperiment man faktisk fikk svar på. Det er en ulempe å bli urettmessig anklaget for UUU-fiske. Samtidig skal det bemerkes at utslag på en eller flere av disse indikatorene alene ikke automatisk betyr at fartøyet driver UUU-fiske. Samtidig vil det ikke direkte registreres personopplysninger, men det er i enkelte tilfeller mulig å identifisere en eier av skip ved å bruke eksterne kilder.

Innsamling og analyse av data gjøres på en lukket plattform. Tilgangen til denne plattform er kun for student og veileder. De innsamlede dataene og resultatene slettes når oppgaven er ferdigstilt.

3 Etterretningsanalyse

Dette kapittelet gir en overordnet beskrivelse av etterretning, etterretningsanalyse, strukturerte analyseteknikker og *«Analysis of competing hypotheses»* (ACH). Kapittelet starter med en kort redegjørelse for hva som ligger i begrepet etterretning og en overordnet gjennomgang av etterretningsprosessen. Deretter fokuseres det på etterretningsanalyse og strukturerte analyseteknikker, før ACH presenteres spesielt. Kapittelet avsluttes med en diskusjon rundt fordeler og ulemper ved ACH som metode. Dette danner grunnlaget for den reviderte metoden som presenteres i kapittel 5.

3.1 Etterretningsteori

Herskere og hærførere har til alle tider kjent til betydningen av etterretning for å nå sine mål. I Det gamle testamentet kan vi lese at Moses sendte spioner til Kanaan, mens de gamle egypterne, grekerne og romerne hadde velutviklede nettverk av spioner. Det å avsløre det som andre ønsker å holde skjult, og benytte dette til egen fordel, har alltid vært sentralt i etterretning (Stenslie et al., 2019, s. 19-21). Etterretning er en viktig del av en stats verktøykasse for å ta gode beslutninger i en usikker og omskiftelig verden.

På tross av denne anerkjennelsen, er det ingen universell, anerkjent etterretningsteori (Warner, 2008). Etterretningsteoretikere er ikke enig om én definisjon av etterretning og debatten dreier blant annet rundt om det er produktet som er etterretning, eller om organisasjonen og aktiviteten som ligger bak, også er det. Her kan man sannsynligvis trekke inn mange ulike forklaringsmodeller, men mye av forklaringen ligger i at forskjellige nasjoner har forskjellige behov og forskjellig kultur (Brantly, 2018). Som et eksempel kan man se på forskjellene mellom hvordan Norge og USA definerer etterretning i en militær kontekst.

Den amerikanske definisjonen legger vekt på at det er produkt, aktivitet og organisasjon:

> 1. The product resulting from the collection, processing, integration, evaluation, analysis, and interpretation of available information concerning foreign nations, hostile or potentially hostile forces or elements, or areas of actual or potential operations. 2. The activities that result in the product. 3. The organizations engaged in such activities (Department Of Defense, 2013, s. GL-8).

Definisjonen i Forsvarets etterretningsdoktrine er noe smalere:

> Etterretning er resultatet av statlig sanksjonert innhenting, analyse og vurdering av data og informasjon, som er generert åpent eller fordekt og utarbeidet for å gi fortrinn i beslutningsprosesser. (Forsvarssjefen, 2021, s. 20)

Vi ser at den norske definisjonen begrenser seg selv til at det handler om aktivitet som foregår i en statlig ramme, og her skiller den seg fra den amerikanske definisjonen. Dette har sitt opphav i etterretningsloven (e-loven) og fokuset på at etterretning er regulert aktivitet. E-loven er utformet med henblikk på at Etterretningstjenesten er en utenlandsetterretningstjeneste som ikke har politioppgaver i sin portefølje, eller utfører oppgaver med polisiære formål (Forsvarssjefen, 2021, s. 8). Dette perspektivet har ikke den amerikanske definisjonen.

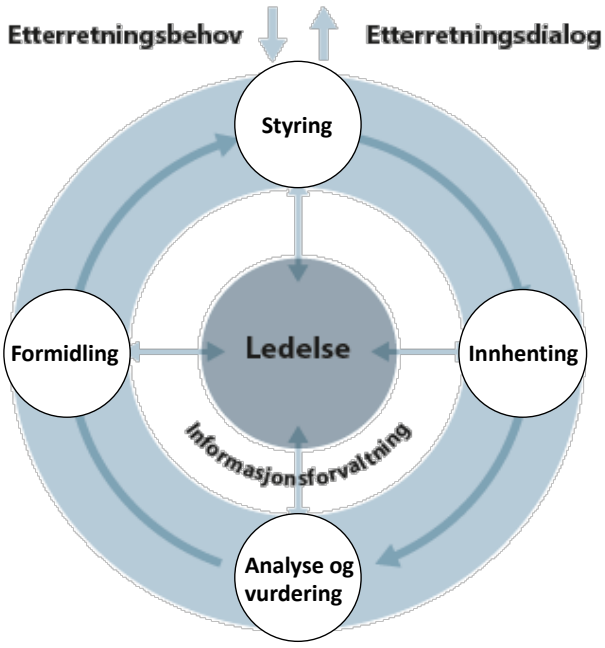
Norske og amerikanske behov er betydelig forskjellige da USA har en rolle som supermakt med globale interesser, kontra Norge med vår betydelig mindre interessesfære. Samtidig ser man også store ulikheter i kultur, åpenhet og spesielt fokus på personvern. Et eksempel som understreker dette er avsløringen av hvordan National Security Agency (NSA) systematisk har innhentet enorme datamengder om et stort antall amerikanske og utenlandske borgere i en årrekke (Edgar, 2017), sammenlignet med den norske offentlige debatten og motstanden mot en mindre inngripende og bedre regulert løsning for lagring av såkalt tilrettelagt innhenting (Forsvarsdepartementet, 2020).

Hvis vi ser på likheter så sier begge definisjonene noe om at dette omhandler både innhenting og vurdering, og man kan tolke det dithen at dette er i den hensikt å kunne utnyttes eller gi fortrinn i beslutningsprosesser. Etterretning er som sådan både et sosialt og politisk fenomen, og ikke kun en teknisk disiplin (Gill & Phythian, 2016, s. 8). Kjernen i etterretningsvirksomhet er derfor å øke beslutningstakernes innsikt og forståelse for å kunne håndtere risiko og fatte bedre beslutninger (Forsvaret, 2019, s. 139; Lim, 2016, s. 623; Omand, 2014).

3.2 Etterretningsprosessen

For å beskrive prosessen som ligger bak et etterretningsprodukt brukes ofte en syklisk prosess som beskriver hva som skjer fra en beslutningstaker formulerer et etterretningsbehov til han mottar et etterretningsprodukt som svarer på hans behov. Prosessen har fire steg: styring, innhenting, analyse og vurdering og formidling (se figur 2), og skal sikre at etterretningsbehovene til kundene/beslutningstakerne besvares. Prosessen omtales som den «sykliske etterretningsprosessen» eller «etterretningshjulet». Ofte vil et etterretningsbehov som er besvart generere nye behov og prosessen visualiseres derfor gjerne som et hjul for å understreke dens sykliske karakter. I virkeligheten er arbeidet i en etterretningsorganisasjon mer komplekst og mange av prosessene vil foregå parallelt. Modellen er generisk, normativ og prinsipiell. Selv om

etterretningsbrukeren, etterretnings-behovene og organiseringen varierer på strategisk, operasjonelt og taktisk nivå, er den prinsipielle beskrivelsen for hvordan etterretning utøves, lik (Forsvarssjefen, 2021, s. 43).



Figur 2: Den sykliske etterretningsprosessen

Prosessen starter med at beslutningstaker har et informasjonsbehov. I dialog med den aktuelle etterretningsenheten verifiseres og omformes dette til et etterretningsbehov. I første delprosess, styring, gjennomgår man det etterretningsgrunnlaget man allerede har, og veier og prioriterer de ulike etterretningsbehov i forhold til de innhentingsressursene man har tilgjengelig. Deretter fordeles innhentingsoppdrag til de ulike innhentingsavdelingene. I andre delprosess, innhenting, hentes det inn informasjon ved hjelp av menneskelige og/eller tekniske innhentingsressurser. I tredje delprosess, analyse og vurdering, systematiseres, analyseres og sammenstilles de innsamlede dataene, og dataene settes i en kontekst. Dette resulterer i etterretningsprodukter. I fjerde steg, formidling, deles produktene med oppdragsgiver. Dette kan være alt fra skriftlig rapporter til muntlige presentasjoner eller diskusjon over kartbordet. Her vil typen oppdrag man understøtter og sjefens preferanser påvirke hvordan produktet leveres.

Selv om prosessen er visualisert i fire delsteg som følger etter hverandre er det i virkeligheten mye større overlapp mellom stegene, og mye dialog på kryss og tvers. En helhetlig ledelse av etterretningsoperasjonene er derfor essensielt for å ivareta en fornuftig ressursbruk og for å ivareta et helhetlig perspektiv. Denne prosessen refereres ofte til som *«Intelligence requirement management collection coordination»* (IRMCC)-prosessen (Etterretningstjenesten, 2013, s. 18). Prosessen er sentral for å koordinere mellom de forskjellige aktørene og delprosessen, samt sørge for at sjefen får besvart sitt informasjonsbehov til rett tid.

3.3 Etterretningsanalyse og strukturerte analyseteknikker

Hensikten med etterretningsanalyse er å trekke de mest korrekte slutningene ut fra den informasjonen som er tilgjengelig. Analytikeren spiller en nøkkelrolle i denne prosessen. Analytikeren

skal vurdere sannhetsverdien og betydningen av innhentede etterretninger, og avgjøre hvor viktige og relevante de er for beslutningstakeren. Den innhentede informasjon må også vurderes mot åpent tilgjengelig informasjon. Ved hjelp av denne prosessen er målet å levere mer presise, relevante og troverdige vurderinger til brukeren (Forsvarssjefen, 2021, s. 67; Omand, 2019).

Etterretningsfaget og spesielt etterretningsanalyse har alltid stått i spagaten mellom håndverk og vitenskap (Brantly, 2018, s. 562), noe som har skapt mye debatt innad i fagfeltet. I essens handler denne debatten om hvorvidt etterretningsanalyse er et håndverk, basert på erfaringer, intuisjon og intuitive vurderinger, eller om det er en vitenskap, som baserer seg på strukturerte, systematiske analysemetoder (Folker, 2000; Stenslie, 2019, s. 66).

Allerede i kjølvannet av andre verdenskrig, fremhevet en av etterretningsprofesjonens nestorer, Sherman Kent, betydningen av å sette vitenskapelige metoder foran intuisjon (Kent, 1949). Kent mente at etterretningsanalyse måtte ta steget fra håndverk til en moderne profesjon hvor utøveren er utdannet innenfor sitt felts vitenskapelige metoder og anvender disse i sitt praktiske virke som analytiker. Imidlertid var det først i 1990-årene at idéene om profesjonalisering og bruk av vitenskapelige metoder for alvor skjøt fart. En løsning på dette ble strukturerte analyseteknikker, innad i etterretningsmiljøet best kjent under akronymet SAT (Stenslie, 2019, s. 67).

SAT er et samlebegrep for en rekke teknikker og prosesser som er utviklet for å forenkle, forbedre og effektivisere etterretningsanalyse, med andre ord hjelpe etterretningsanalytikeren med å gjøre bedre vurderinger. SAT kan benyttes i hele eller deler av prosessen, som enkeltteknikker, et sett av teknikker, eller som et supplement (Forsvarssjefen, 2021, s. 67). Teknikkene er i stor grad utviklet av de tidligere CIA-analytikerne Richards J. Heuer Jr. og Randolph H. Pherson. De har sammen gitt ut boka *«Structured Analytic Techniques for Intelligence Analysis»* (Heuer & Pherson, 2010), som i dag er pensum på de fleste innføringskurs i SAT (Stenslie, 2019, s. 65-74).

Etterretningsanalyse er først og fremst en kognitiv prosess og forskning viser at slike prosesser ofte er utsatt for en rekke systematiske skjevheter. Disse skjevhetene har sitt opphav i hvordan de kognitive prosessene blir behandlet i hjernen. I følge Kahneman (2011) vil de kognitive prosessene, avhengig av problemet vi står overfor, enten bli utført intuitivt, såkalt system 1-tenkning, eller ved resonnement, såkalt system 2-tenkning. Mens system 1 virker automatisk og hurtig, med liten eller ingen anstrengelse og ingen opplevelse av viljekontroll, vil system 2 tildele oppmerksomhet til de anstrengende mentale aktivitetene som krever det. Arbeidsdelingen mellom system 1 og system 2 er meget effektiv og er optimalisert for høyest mulig ytelse med minst mulig anstrengelse. Denne

arbeidsdelingen fungerer for det meste meget godt fordi system 1 er veldig effektivt. Utfordringen ligger i at hjernen vår foretrekker kognitiv letthet og foretrekker å benytte system 1, fremfor det mer anstrengende system 2. Hjernen skaper derfor intuitivt mønstre, relasjoner, sammenheng og logikk, selv når det ikke er noen, ofte basert på svært lite data eller informasjon. Disse forenklede mentale modellene av virkeligheten lagres i langtidshukommelsen for å tas fram ved senere anledninger og til bruk ved fremtidige beslutninger. Dette gir rom for mentale fallgruver og logiske feilslutninger.

Et eksempel på en slik fallgruve som etterretningsanalytikere ofte er utsatt for er bekreftelsesbias eller bekreftelsestendens (Heuer, 1999). Denne kan gjøre seg gjeldende både når informasjon skal hentes inn og når informasjon skal tolkes eller analyseres. Når vi henter inn informasjon har vi en tendens til å legge merke til eller søke etter det som bekrefter noe man tror. Når vi analyserer har vi en tendens til å legge vekt på informasjon som stemmer med hva vi tror på forhånd, dette kan medføre at ellers nøytral informasjon får en overdreven betydning (Svartdal, 2019). Slik tenkning gjør oss sårbare for feil fordi vi overser og feiltolker informasjon som tilsier at vi burde endre oppfatning. Dette synet støttes også av Popper (2002, s. 47) som poengterte at vi kan verifisere eller bekrefte en hver teori hvis vi leter etter bekreftelse. Det er derfor et ideal å falsifisere hypoteser fremfor å bekrefte, noe som også benyttes i flere av de strukturerte analyseteknikkene.

Oppsummert kan man derfor si at SAT forsøker å bidra til å redusere slutningsfeil og redusere sannsynligheten for å ta mentale snarveier, og gjennom dette styrke etterretningsproduktenes pålitelighet (Forsvarssjefen, 2021, s. 67-69; Heuer, 1999).

3.4 Analysis of competing hypotheses

Analysis of competing hypotheses (ACH) er en metode for å monitorere og vurdere sannsynligheten av ulike hypoteser. Metoden ble utviklet på 1970-tallet av Richards Heuer med bakgrunn i hans «*never-ending quest for better analysis»* (Heuer, 1999). ACH er blant de mest populære og underviste innen strukturerte analyseteknikker og er en av få diagnostiske teknikker som nevnes i CIAs *«Tradecraft Primer: Structured Analytic Techniques for Improving Intelligence Analysis»* (Central Intelligence Agency, 2009). Her bemerkes det at metoden *"has proved to be a highly effective technique when there is a large amount of data to absorb and evaluate"* (Jones, 2018, s. 273). Metoden er også å finne igjen i «*UK Ministry of Defence's Quick Wins for Busy Analysts Handbook»* (UK Ministry of Defence, 2013).

Metoden er utviklet for å være en effektiv og pålitelig metode for å teste konkurrerende hypoteser. Samtidig evaluering av ulike, konkurrerende hypoteser er vanskelig uten analytiske hjelpemidler, og

for å håndtere denne kompleksiteten legger Heuers modell opp til bruken av analytiske hjelpemidler, samt utstrakt eksternalisering av analyseprosessen. Poppers (2002) teorier om at hypoteser bør avkreftes er sentralt i metoden og metoden fokuser på å avkrefte istedenfor å bekrefte hypoteser (Heuer & Pherson, 2010, s. 160). Heuer erkjenner at metoden ikke nødvendigvis alltid gir det rette svaret, men det gir analytikeren noen verktøy for å redusere sannsynligheten for feilslutninger (Heuer, 1999).

ACH steg for steg

ACH er anvendelig når man har et fenomen som har flere mulige forklaringer. Fenomenet kan ha allerede ha skjedd eller er forventet å skje en gang i fremtiden. Resultatet av prosessen er at man står igjen med en rangering om hvorvidt en eller fler forklaringer er mer sannsynlig enn andre.

Nedenfor er metoden beskrevet stegvis slik den er presentert i *«Structured Analytic Techniques»* (Heuer & Pherson, 2010, s. 162-164):

Steg 1: Utvikle hypoteser. Det første steget er å identifisere de aktuelle hypotesene. Det er sentralt at hypotesene dekker flere mulige forklaringer slik at mulighetsrommet spennes ut. Hypotesene trenger ikke dekke alle muligheter, men må favne om alle *rimelige* muligheter. For hver hypotese bør det lages et narrativ og hypotesene bør være gjensidig utelukkende.

Steg 2: List opp signifikante bevis. List opp alle bevis, argumenter eller antagelser. Det er også interessant å liste opp indikatorer som man *ikke* vil se, gitt at hypotesen er sann. Bevis som både bekrefter og avkrefter hver av hypotesene bør inkluderes.

Steg 3: Vurder diagnostisk score. Listen med signifikante bevis identifisert i steg 2 må vurderes med tanke på den diagnostiske scoren. For å vurdere dem opp mot hverandre settes alle bevis, argumenter eller antagelser opp i en matrise og vurderes opp mot alle hypotesene. Hensikten er å fjerne bevis med lav diagnostisk score, altså de indikatorene som hverken bekrefter eller avkrefter noen av hypotesene.

Steg 4: Revurder hypotesene. Vurder om noen av hypotesene kan kombineres eller om nye hypoteser må legges til. Hvis nye hypoteser legges til gjennomføres steg 2 og 3 på nytt.

Steg 5: Skisser foreløpige konklusjoner om den relative sannsynligheten for hver hypotese. Dette gjøres ved å summere inkonsistens-scoren til hver hypotese. Deretter rangeres hypotesene slik at den høyest rangerte er den med lavest inkonsistens-score.

Steg 6: Analyser robustheten til den foreløpige konklusjonen gitt endringer i tolkningen i noen av de sentrale bevisene. Hvis en eller flere av disse informasjonsbitene var feil, villedende eller tolket på en annen måte, vil dette endre konklusjonen? Hvis ja, gå tilbake og dobbeltsjekk tolkningen av disse bevisene.

Steg 7: Rapporter slutningene til beslutningstaker eller kolleger. I framleggingen er det viktig å diskutere den relative sannsynligheten for alle hypotesene og poengtere sentrale bevis. Forklar også hvorfor alternative hypoteser ble avvist.

Steg 8: Utvikle indikatorer. Lag to lister — en med fokus på fremtidige hendelser som vil bekrefte den nåværende konklusjonen, og en liste med fokus på hendelser og informasjon som avkrefter konklusjonen. Begge listene monitoreres jevnlig og brukes til å vurdere om nye bevis enten styrker eller svekker slutningene.

Terrorism in Tokyo From Aum Shinrikyo

		Weight	H: 1	H: 4	H: 2	H: 3
			Kooky Cult	Terrorist Group	Political Movement	Criminal Group
	Inconsistency Score ⇨		-1.0	-1.0	-2.0	-3.0
E3	Attacks on Journalists	MEDIUM	I	N	I	I
E2	Religious Affiliation	MEDIUM	C	I	I	I
E4	Established Party	MEDIUM	N	N	C	I
E1	Blind Leader Mastsumoto	MEDIUM	C	C	C	C

Figur 3: Eksempel på utfylt ACH-matrise (Central Intelligence Agency, 2009, s. 15)

Eksempelet i figur 3 er hentet fra *«Tradecraft Primer: Structured Analytic Techniques for Improving Intelligence Analysis»* (Central Intelligence Agency, 2009) og viser bruken av ACH ifm et terrorangrep i Tokyo i 1995. Her er problemstillingen knyttet til hvem eller hvilken gruppe som utførte angrepet. I eksempelet finner vi hypotesene (H1 til H4), hvilke bevis som er listet opp (E1 til E4), hvorvidt bevisene er konsistente (C) eller inkonsistente (I) med hypotesene og total inkonsistens score. Basert på denne matrisen er H1 og H4 i minst grad avkreftet og er derfor de mest sannsynlige hypotesen. Det er også verdt å merke seg «E1 – Blind Leader Mastsumoto» som er et eksempel på et bevis med lav diagnostisk score da dette beviset bekrefter alle hypotesene.

3.5 Hvor effektive er ACH og SAT?

Skaperen av metoden Richards Heuer trekker, ikke overraskende, fram flere fordeler med metoden. For det første tvinger den analytikeren til å tidlig utvikle et full sett med alternative hypoteser, noe som reduserer sannsynligheten for *«satisficing»,* det vil si at man går for den første og beste løsningen som dukker opp. Videre hjelper den analytikerne til å lete etter bevis som avkrefter hypotesene, ikke kun bekrefter dem. For det tredje hjelper metoden med å sortere bevis på en analytisk måte. Avslutningsvis argumenterer han for at metoden øker transparensen i slutningene da det er tydelig hvordan det enkelte bevis vektes (Heuer & Pherson, 2010, s. 161-162).

Folker (2000) sin studie viste at analytikere som benyttet ACH var mer objektive i slike analyser sammenlignet med analytikere som benyttet mer intuitive metoder. En annen studie viste at ACH drastisk reduserte bekreftelsestendensen, samtidig som det medførte at analytikerne søkte etter og brukte tre ganger mer informasjon, sammenlignet med kontrollgruppen (Brasfield, 2009). Coulthart (2016) på sin side fant at effektiviteten til strukturerte analyseteknikker generelt var overvurdert, med unntak av ACH og to andre metoder.

Selv om ACH har blitt de facto standard for hypotesetesting, spesielt i vestlige etterretningstjenester, er det visse utfordringer ved å bevise effektiviteten til metoden. Dhami et al. (2019) fant i sin studie ut at analytikere som var trent i ACH ikke gjorde det noe bedre enn analytikere som benyttet intuitive metoder. Et av deres sentrale funn er at selv trente og instruerte analytikere ikke følger metoden og avviker fra et eller flere steg underveis. Samme studie viser også til at de finnes en rekke andre utfordringer; det er uklart hvordan hypoteser skal velges; det er uklart hvilke kriterier som skal brukes for å vurdere bevis som samsvarende eller inkonsekvent med en hypotese; og det er uklart hvilke kriterier som skal brukes for å vurdere diagnostisitet. Dhami poengterer at han er overrasket over populariteten til metoden da mye av forskningen på ACH viser at metoden ikke er signifikant bedre enn intuitive metoder (Dhami et al., 2019, s. 1-2).

Mandel (2020) går enda hardere til verks i sin artikkel hvor han hevder at de ikke finnes vitenskapelige studier som viser at ACH faktisk fungerer. Videre sår han tvil om hele grunnlaget til SAT da han mener at grunnlaget det bygger på er for svakt. Hans kjernepoeng er at Heuer sine teknikker er utdaterte og modne for oppdateringer, og at etterretningsanalytikere generelt fortjener bedre metoder.

Fra egen erfaring er en av de største fordelene med SAT generelt og ACH spesielt at det tvinger fram en eksternalisering av prosessen. Denne eksternaliseringen er til stor hjelp når man skal jobbe

sammen i team. Dagens operasjonsmiljø er mer komplekst enn tidligere, og man må jobbe i team bestående av ulik fagekspertise, for å kunne gjøre gode analyser (Borg, 2017). For slike gruppeprosesser er det sentralt at premisser og delslutninger som analysene hviler på er godt kjent innad i teamet. Dette skaper en transparens i prosessen som er viktig for å kunne utnytte synergiene og skape en helhet i analyseprosessen. Dette underbygges også av funn i Stenslie (2019, s. 75-76) sin SAT-undersøkelse hvor denne transparensen nevnes som en av fordelene analytikere og ledere ser ved å benytte SAT. Heuer poengterer også denne eksternaliseringen i sin introduksjon av metoden (Heuer & Pherson, 2010, s. 161). Det er også verdt å merke seg at de studiene som har testet effektiviteten av ACH har testet enkeltindivider og målt den enkeltes prestasjoner. Det er ikke gjort studier som har målt effekten av SAT på gruppeprosesser og samarbeid om komplekse problemstillinger.

ACH har sine sterke og svake sider, og kritikken mot metoden er ganske krass. Samtidig er det heller ingen som har kommet opp med alternative metoder som er bevist å være bedre. SAT generelt og ACH spesielt vurderes derfor fortsatt til å være relevante metoder og verktøy som kan hjelpe analytikere med å unngå etterretningsfeil, men metodene har sine svakheter som man må være bevisst.

4 Stordata

Dette kapittelet gir en overordnet beskrivelse av stordata og stordataanalyse. Kapittelet starter med en kort redegjørelse for hva som ligger i begrepet stordata, og presenterer grunnleggende og sentrale begreper som er relevante for denne oppgaven. Deretter redegjøres det for strømmesystemer og stordataanalyse. Kapittelet gir et grunnlag for å forstå endel sentrale begreper innen stordata og dette danner grunnlaget for den reviderte metoden som presenteres i kapittel 5.

4.1 Hva er stordata?

Stordata som begrep kan spores tilbake til 1990-tallet (Balazka & Rodighiero, 2020, s. 2) og diskusjoner om forvaltning av store mengder data og datasett i både akademia og industrien (Vivento & Kaupang, 2015). Den første formelle akademiske definisjonen kom allerede i 2000, men fenomenet har først og fremst hatt sin fremvekst siden 2012 (Balazka & Rodighiero, 2020, s. 2).

I litteraturen om stordata verserer det en rekke ulike definisjoner og akademia sliter med å enes om en felles definisjon. Gartner (2012) presenter en definisjon hvor de knytter det til teknologien, men poengterer at det må gi merverdi: «*Stordata er høy-volum, høy-hastighet, og/eller høy-varierte informasjonsressurser som krever nye former for prosessering for å kunne understøtte bedre beslutninger, innsikt og prosessoptimalisering*.»

Vivento & Kaupang (2015) går enda bredere i sin definisjon og inkluderer også et perspektiv på kvalitet og merverdi:

> Stordata er analyse av massive samlinger av data, med stor variasjon i datakilder og formater, hvor datasettet oppdateres med høy frekvens og hvor grunnlagsdataenes opprinnelse og kvalitet er avklart og hvor analysen av datasamlingene gir økt verdi i forhold til hva datakildene enkeltvis ga. (Vivento & Kaupang, 2015).

Stolpe et al. (2019, s. 8) omtaler stordata mer som en betegnelse på en samfunnsutvikling enn en spesifikk teknologi og sier at det også kan forstås som et begrep som omfatter en rekke informasjonsbehandlingsproblemer. De poengterer videre at stordata må forstås som en samlebetegnelse for beregningsoppgaver som enten er for komplekse eller som vokser for raskt til at de kan håndteres av en enkelt maskin. Det er verdt å merke seg at det her fokuseres på manipulering og analyse av disse dataene.

Stordata handler altså om utnyttelse av store mengder data, gjerne på tvers av virksomheter, datakilder og formater. Stordata er et sett med teknologier og omfatter hele verdikjeden. Dette omfatter datainnsamling, lagring, prosessering, analyse og visualisering av resultater (Vivento &

Kaupang, 2015). Med andre ord dreier stordata seg ikke nødvendigvis bare om datamengder, men også om opprinnelse, betydning, formater og hastigheter (Stolpe et al., 2019).

Data har til alle tider blitt benyttet som grunnlag for å fatte beslutninger og gir i seg selv ingen mening eller verdi, om det ikke gjøres en tolkning eller analyse slik at dataene settes i en eller annen kontekst. Først da kan data bli til informasjon og kunnskap. I dette ligger en påminnelse om at stordata først har verdi når de blir utnyttet.

4.2 De tre V'ene

Selv om det er uenighet om definisjonen av stordata, virker det å finnes et minste felles multiplum de fleste kan si seg enige i. Dette er en liste med egenskaper referert til som «De tre V'ene» som ble utarbeidet av Laney (2001) på tidlig 2000-tall.

Volume = volum: Store mengder av data, fra datasett med størrelser fra terrabytes til zettabytes. Størrelsen på et datasett er det mest opplagte kjennetegnet ved et stordataproblem. Dersom mer enn en maskin er nødvendig for å lagre og behandle et datasett, så er datasettet for alle praktiske formål «stort». Per definisjon henger derfor stordata nært sammen med parallell og distribuert databehandling. Siden behovet for regnekraft og minne overgår det en enkelt maskin kan tilby, dreier stordata seg i stor grad om å koordinere ressurser over en klynge (cluster) av maskiner. Algoritmer som er i stand til å bryte opp en beregningsoppgave i mindre, uavhengige delproblemer er en avgjørende del av dette (Stolpe et al., 2019, s. 9).

Velocity = omløpshastighet: Store mengder data fra transaksjoner med høy oppdateringsfrekvens resulterer i datastrømmer som kommer i stor hastighet. Tiden til rådighet for å handle på bakgrunn av slike datastrømmer vil ofte være kort (Vivento & Kaupang, 2015). Typiske eksempler på slike systemer vil være systemer for hendelsesdeteksjon og bildebygging. I slike systemer flyter gjerne informasjonen inn fra flere kilder samtidig, og må analyseres fortløpende for at forståelse av den gitte situasjonen skal holde seg så fersk som mulig. Dette handler altså ikke om lagringskapasitet, men om det man kan kalle evnen til å absorbere en strøm (Stolpe et al., 2019, s. 9).

Variety = variasjon: Data kommer fra ulike datakilder – både fra eksterne kilder og fra datakilder internt i virksomheten som skal bruke dem. Men dataene har også variabel form; transaksjons- og loggdata fra ulike applikasjoner, strukturerte data som rader av data, ustrukturerte data som tekst, bilder, videostrømmer, lyd m.m (Vivento & Kaupang, 2015). Et stordatasystem må være i stand til å utvinne verdifull informasjon ved å konsolidere slike svært heterogene kilder – dvs. ved å omforene ulike typer data (Stolpe et al., 2019, s. 9).

4.3 Stordatainfrastruktur

For å løse et stordataproblem er man avhengig av en stordatainfrastruktur. Det vil si en rekke programmer eller applikasjoner som i samarbeid løser stordataproblemet. Håndteringen av store datasett kan være svært ressurskrevende og krever utstyr og programvare som sikrer tilstrekkelig datakvalitet. Man må også disponere tilstrekkelig transport- og lagringskapasitet samt regnekraft for å kunne bearbeide og analysere dataene. Utbygging og utvikling av slike ressurser kan være både kostnadskrevende og teknologisk utfordrende (Dvergsdal & Elster, 2019).

Det ikke én generisk stordatainfrastruktur som passer alle brukstilfeller. Hvilke komponenter infrastrukturen består av bør dikteres av de karakteristiske trekkene ved det problemet som skal løses (Halvorsen & Hansen, 2020, s. 4) og valget av programvarepakke bør være basert på en inngående analyse og forståelse av hvilket problem det er man forsøker å løse. Hva slags data dreier det seg om, komplekse eller enkle, homogene eller heterogene? Hva slags algoritmer skal brukes og hvor oppdaterte trenger dataene å være? Med andre ord, å velge rett programvare for et bestemt stordataproblem er ikke en triviell oppgave, men krever at man på bakgrunn av en forståelse av problemet har en noenlunde klar tanke om hvilke egenskaper man ønsker at systemet skal ha. (Stolpe et al., 2019, s. 12).

Normen i dag er å bygge stordatasystemer basert på programvare med åpen kildekode. Dette gir gode muligheter til å sette sammen komponenter etter behov, og prøve ut ulike komponenter etter prøv-og-feil-prinsippet, da det stort sett ikke er forbundet en direkte kostnad å teste og bruke slik programvare. Det finnes et rikt utvalg av komponenter som kan settes sammen, noe som gjør at design av stordatainfrastruktur i stor grad kan betraktes som skreddersøm. Stolpe et al. (2019) vurderer også at tilgangen på moden programvare er såpass god at det ikke er nødvendig i dag å kjøpe dyre proprietære systemer der man låser seg til en leverandør.

4.4 Strømmesystemer

Stordatasystemer kan grupperes etter hvilke datastrukturer de er bygget over. Med en hensiktsmessig grovsortering dreier det seg om de følgende tre typene:

- Systemer som representerer data i *tabeller*
- Systemer som representerer data i *grafer*
- Systemer som beregner *datastrømmer*

Videre vil det kun redegjøres for strømmesystemer da det er denne typen systemer som er relevante for denne oppgaven. For en mer utfyllende beskrivelse av systemer som behandler graf og tabulære data henvises det til litteraturen (Stolpe et al., 2019).

Verden flommer i dag over av data, og mange av disse dataene kjennetegnes av at de er strømmende. En strømmende datakilde gir en kontinuerlig og vilkårlig lang sekvens av data. Eksempler på slike datakilder er AIS-data, video-overvåkningssystemer og datastrømmer fra sosiale medier.

Strømmende datakilder har potensiale til å bidra med ny informasjon som kan hjelpe oss å bygge opp en forståelse av situasjonen rundt oss, men for å utnytte dette potensialet må vi være i stand til å behandle disse dataene raskt nok til at informasjonen som kan utledes fortsatt er nyttig. Den største fordelen ved å behandle data som en strøm, er at analyse kan gjøres fortløpende og resultater kan leveres raskt. For å klare dette, holder og prosesserer strømmesystemer typisk data og resultater i minnet. På den måten unngås den tidkrevende lesingen og skrivingen av data og resultater til disk. Dette betyr imidlertid at strømmesystemer stiller store krav til minne i infrastrukturen, og at med mindre man også har en prosess som også skriver data til disk (persisterer) vil dataene i et strømmesystem gå tapt når systemet slås av. De fleste strømmesystemer har imidlertid mulighet for å lagre de strømmende dataene (Stolpe et al., 2019, s. 50-53).

4.5 Stordataanalyse

Å oppdage mønstre, foreslå konklusjoner og støtte beslutningsprosesser, eller sagt på en annen måte å foredle rådata til nyttig informasjon er målsetninger for de aller fleste anvendelser av stordataanalyse. I denne oppgaven benyttes det flere ulike analyseteknikker, men de faller inn under kategorien hendelsesbehandling (*event processing).* Dette er et sett av teknikker for å identifisere og behandle hendelser i en kontinuerlig strøm av sanntids- eller nær sanntidsdata. Hensikten er å kunne skille ut gitte situasjoner, eller kjenne igjen mønster, i den gitte datastrømmen. Hendelsesdeteksjon omfatter statistiske og logikkbaserte teknikker (Stolpe et al., 2019, s. 12-13). En logikkbasert teknikk kjennetegnes ved at du forklarer maskinen hva den skal se etter, til forskjell fra statistikkbaserte teknikker hvor maskinen i større grad finner mønster basert på statistikk. Denne oppgaven fokuserer på logikkbaserte teknikker.

5 Stordatadrevet ACH

I dette kapittelet vil en metodikk for å kombinere etterretningsanalyse og stordataanalyse bli presentert. Metoden blir benevnt «Stordatadrevet ACH» og er nyskapningen i denne oppgaven. Først vil den nye metoden bli presentert steg for steg, deretter vil metoden sammenlignes med Heuers opprinnelige ACH-metode for å tydeliggjøre likheter og forskjeller i metodene og hvor den nye metoden avviker fra originalen.

5.1 Stordatadrevet ACH

Den reviderte metoden har som ambisjon å forene de sterke sidene til menneske og datamaskin. Datamaskiner er overlegne oss mennesker når det gjelder å behandle store informasjonsmengder raskt og med et pålitelig resultat, men datamaskiner har ikke den kreativiteten og evnen til å utvikle hypoteser, se helhet og vurdere kontekst som mennesker har.

I stordatadrevet ACH vil maskinens funksjon være å hjelpe etterretningsanalytikeren å håndtere store informasjonsmengder, trekke ut det som er relevant og presentere denne informasjonen i strukturert form. Etterretningsanalytikeren vil utvikle hypoteser, utforme indikatorer og algoritmer, og ta de endelige beslutningene.

Som utgangspunkt for å bruke metoden må man ha en *problemstilling* som skal løses, og som er hensiktsmessig å løse med stordatadrevet ACH. Det sentrale poenget her er å analysere datagrunnlaget og vurdere om dette er av en sånn karakter at det er egnet til å bruke som datakilde, og om datavolumet er stort nok til at man kan dra nytte av metoden. En slik analyse bør foregå i tett samarbeid mellom analytiker og dataingeniør for å sikre at alle perspektiver er belyst og analysert. Når denne analysen er gjort kan man starte med steg 1 i metoden. Steg 1 til 4 vil normalt etterretningsanalytiker(e) gjennomføre uten støtte fra dataingeniør(er), mens steg 5 til og 8 bør gjøres som et samarbeid mellom etterretningsanalytiker og dataingeniør. Det er verdt å bemerke at det ikke er noen ulempe om hele prosessen gjennomføres i fellesskap. Nedenfor presenteres metoden stegvis.

Steg 1: Utvikle hypoteser. Det første steget er å identifisere de aktuelle hypotesene. Det er sentralt at hypotesene dekker flere mulige forklaringer slik at mulighetsrommet spennes ut. Hypotesene trenger ikke dekke alle muligheter, men må favne om alle *rimelige* muligheter. For hver hypotese bør det lages et narrativ og hypotesene bør være gjensidig utelukkende.

Steg 2: List opp signifikante bevis. List opp alle bevis, argumenter eller antagelser. Det er også interessant å liste opp indikatorer som man *ikke* vil se, gitt at hypotesen er sann. Bevis som både bekrefter og avkrefter hver av hypotesene bør inkluderes.

Steg 3: Vurder diagnostisk score. Listen med signifikante bevis identifisert i steg 2 må vurderes med tanke på den diagnostiske scoren. For å vurdere dem opp mot hverandre settes alle bevis, argumenter eller antagelser opp i en matrise og vurderes opp mot alle hypotesene. Hensikten er å fjerne bevis med lav diagnostisk score, altså de indikatorene som hverken bekrefter eller avkrefter noen av hypotesene.

Steg 4: Revurder hypotesene. Vurder om noen av hypotesene kan kombineres eller om nye hypoteser må legges til. Hvis nye hypoteser legges til gjennomføres steg 2 og 3 på nytt.

Steg 5 - Velg stordatainfrastruktur. Datastrømmens egenart og hvilke krav som stilles til løsningen vil diktere hvilken stordatainfrastruktur som velges. Valget må være tuftet på en inngående analyse, og en tett dialog mellom etterretningsanalytiker og dataingeniør for å sikre at løsningen gjør det den skal og ivaretar analytikerens behov. Løsningen som velges bør være slik at den relativt enkelt kan integreres i verktøyene som analytikeren bruker i det daglige.

Steg 6 – Bryt ned indikatorer. De indikatorene som har høy diagnostisk score fra steg 3 må deretter brytes ned slik at de er lesbare av den valgte programvaren. Hvordan indikatorene dekomponeres dikteres av datagrunnlag, indikatorens egenart og evt. programmeringsspråk. Her vil valg av stordatainfrastruktur, programmeringsspråk og analytikerens kompetanse diktere hvordan arbeidsfordeling blir mellom dataingeniør og analytiker. Det er en fordel, men ikke et krav, at analytikerne har en grunnleggende forståelse for programmering slik at disse enklere kan skjønne hvordan en indikator må beskrives for å være maskinlesbar.

Steg 7 – Hent inn data. Når løsningen er satt opp og konfigurert starter innhentingsfasen. Stordatainfrastrukturen vil starte nedlasting av de ulike stordatakildene og analyseprogramvaren vil ved hjelp av algoritmene programmert i steg 6 søke etter indikatorene. Lengden av innhentingsfasen styres av tid tilgjengelig, datatilgang og problemets egenart.

Steg 8 – Presenter funnene. Når algoritmene får treff på en indikator presenteres dette for analytikeren. Dette kan gjøres på et utall måter, f.eks. kan hvert eneste funn presenteres, eller når man har et antall treff som er høyere enn en gitt terskelverdi. Det kan også skilles på hvor viktig hver indikator er, og funn kan presenteres etter viktighet. Funnene kan presenteres i matriseform med de

ulike bevisene uthevet, og en eventuell totalscore. Analytikeren må deretter vurdere bevisene og si seg enig eller uenig i funnene, og beslutte om dette skal rapporteres videre eller ikke. Hvis ikke funnene er gode nok må det vurderes om indikatorer og/eller søkemetodikk skal justeres og steg 6 gjentas.

Steg 9 – Formidle: Det siste steget er å trekke endelige konklusjoner og formidle dette videre til beslutningstaker i ønsket format. Form og farge på en slik presentasjon vil avhenge av lokale forhold.

5.2 Sammenligning av Heuers ACH og stordatadrevet ACH

Figuren under viser en sammenligning av Heuers opprinnelige ACH metode og den foreslåtte presentert i kapittel 5.1. Metodene er identiske frem til og med steg 4, men er ulike fra steg 5 og utover.

Heuer ACH	Stordatadrevet ACH
Steg 1: Utvikle hypoteser	Steg 1: Utvikle hypoteser
Steg 2: List opp signifikante bevis	Steg 2: List opp signifikante bevis
Steg 3: Vurder diagnostisk score	Steg 3: Vurder diagnostisk score
Steg 4: Revurder hypotesene	Steg 4: Revurder hypotesene
Steg 5: Skisser foreløpige konklusjoner	Steg 5: Velg stordatainfrastruktur
Steg 6: Analyser robustheten	Steg 6: Bryt ned indikatorer
Steg 7: Rapporter slutningene	Steg 7: Hent inn data
Steg 8: Utvikle indikatorer	Steg 8: Presenter funnene
	Steg 9: Formidle

Figur 4: Sammenligning av Heuers ACH og stordatadrevet ACH

Hovedforskjellen mellom metodene er at i stordatadrevet ACH prioriteres integrering av stordatainfrastruktur og nedbryting av indikatorer, fremfor en grundigere gjennomgang og verifisering av bevisene. I tillegg har stordatadrevet ACH en eksplisitt innhentingsfase (steg 7). Den opprinnelige metoden til Heuer er i noe større grad iterativ da det siste steget er å identifisere nye indikatorer. I Heuers steg 8 skal man *«Identifisere fremtidige indikatorer»* og resultatet av dette steget vil være indikatorer som innhentingsenhetene i en etterretningsorganisasjon aktivt kan lete etter. Som resultat av innhentingen kan man da få ny informasjon som enten avkrefter eller

bekrefter noen av hypotesene. Ny informasjon vil gjøre at prosessen starter på nytt fra steg 2, og man vil vekte den nye informasjon opp mot det informasjonsgrunnlaget man allerede har.

Stordatadrevet ACH vil ikke ha en tilsvarende prosess når det kommer inn ny informasjon, da indikatorene man leter er definert før innhentingen starter. I tillegg er det også vurdert hvordan disse svarene vil påvirke hypotesene. Dette gjør at man ikke kan vurdere noe annet enn det man allerede har definert som indikatorer. Allikevel er det iterative elementer da det alltid vil være behov for å revidere indikatorene ettersom man får ny kunnskap eller når målene man observerer endrer modus operandi.

6 Valg av stordatainfrastruktur

For å evaluere metoden presentert i kapittel 5 ble det gjennomført et felteksperiment. I et felteksperiment testes artefaktet ut under normale betingelser i sine tiltenkte omgivelser (Stølen, 2019, s. 94-95). For å gjennomføre et eksperiment som er i tråd med dette må det etableres en stordatainfrastruktur.

I dette kapittelet vil derfor valgt stordatainfrastruktur, både software og hardware, presenteres. Innledningsvis vil det redegjøres for hvordan en dataflyt må behandles for å at man skal nå målsetningen om å gi dataene merverdi. Deretter vil de ulike komponentene i stordatainfrastrukturen redegjøres for, før det avslutningsvis forklares hvordan dataflyten går gjennom den valgte stordatainfrastrukturen.

6.1 Fra data til merverdi

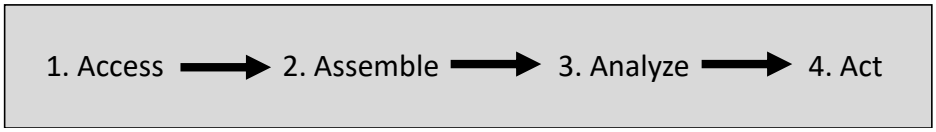
Stordata i seg selv gir ingen mening og dataene må derfor behandles eller manipuleres slik at det kan trekkes ut eller skapes en merverdi av datagrunnlaget. For å skape denne merverdien må dataene behandles i en rekke ulike steg. En modell for å forklare disse stegene fra datakilde til sluttbruker er «4A» (se figur 5). De fire A'ene er *Access, Assemble, Analyze, Act* (Emani, Cullot & Nicolle, 2015).

Access: Stordatakilde må velges og tilgang til datakilden må skaffes. Infrastrukturen må håndtere innhenting fra ulike kilder med ulike protokoller inn i infrastrukturen.

Assemble: Stordatainfrastrukturen er nødt til å håndtere ulike dataformater og kunne trekke ut de ønskede attributtene. I dette steget skal dataene forberedes for analyse og lagres på ønsket sted. Denne prosessen refereres også til som Extract, Transform, Load (ETL).

Analyze: I dette steget foregår selve analysen og her benyttes det spørringer, algoritmer og modeller for å genere ny kunnskap ut av datagrunnlaget.

Act: Avslutningsvis skal dataene presenteres for sluttbruker på en slik måte at de gir beslutningsstøtte. Det er her essensielt at sluttbruker får presentert dette på en slik måte at det er forståelig og muliggjør bedre beslutninger.



Figur 5: Rammeverket 4A

6.2 Software-komponenter

Valg av software-komponenter benyttet i eksperimentet baserer seg på en analyse av datasettet og problemstillingen eksperimentet skal løse. I tillegg bygger den også på «best practice» i stordataanalyse (Halvorsen & Hansen, 2020; Stolpe, Hansen, Halvorsen & Opland, 2020).

Apache Kafka

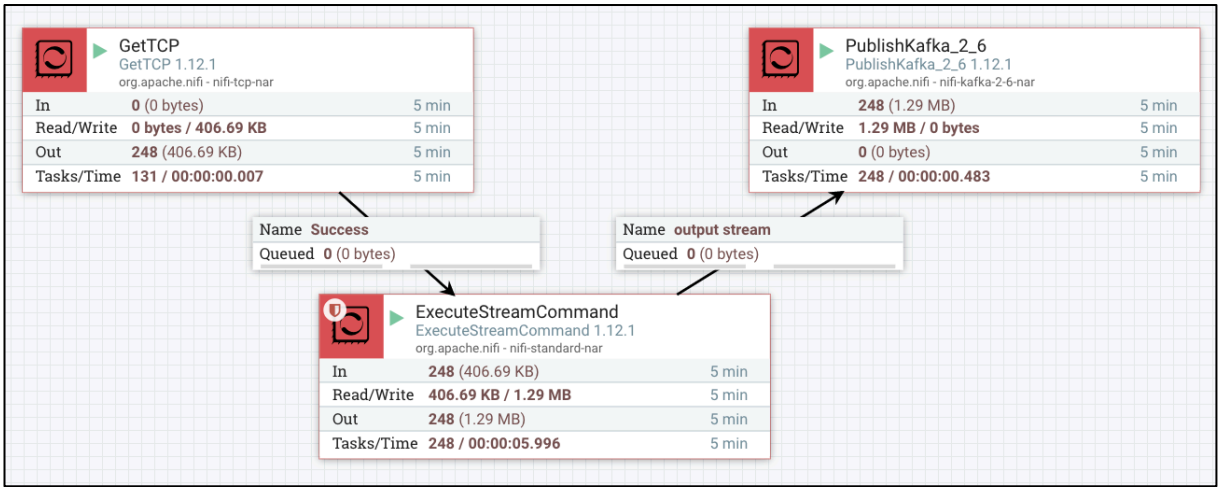
Kjernen i løsningen er Apache Kafka. I tilknytning til denne kjernen er det knyttet en rekke applikasjoner som løser mer eller mindre spesialiserte oppgaver. Dataene flyter mellom komponentene ved hjelp av ulike APIer og plug-iner.

Apache Kafka er et distribuert meldingsverktøy som organiserer innkommende data i merkede køer (*topics*) slik at klienter kan tegne abonnement på data de er interessert i og motta disse når de ankommer Kafka. Dataene som lagres er uforanderlige (*immutable*). Det betyr at alle dataelementer som går gjennom Kafka lagres og aldri endres. Som andre strømmesystemer holder Kafka alle dataene i minnet og krever derfor mye minne, men data kan også lagres til disk (persisteres). Kafka er først og fremst en meldingsbuss som kan fange de strømmende dataene, lagre disse og gi dem videre til andre komponenter som skal stå for analyse og videre prosessering. Ved behov for reprosessering av resultater hentes bare de gamle dataene fram igjen (Stolpe et al., 2019, s. 54).

Apache NiFi

Apache NiFi benyttes for innmating og preprosessering av dataene. NiFi er et program hvor en setter opp ulike prosessorer til innhenting, prosessering, og videresending av data. Dette tilsvarer ETL som beskrevet ovenfor. Figur 6 viser et eksempel på en slik prosess som er satt opp ifm. gjennomføringen av eksperimentet. Her hentes AIS-strømmen inn i NIFI ved å laste ned data fra en gitt IP-adresse vha. funksjonen *«GetTCP»*. AIS-dataene overføres i et binært format og må derfor konverteres før de kan lastes inn i Kafka. Ved hjelp av funksjonen *«ExecuteStreamCommand»* og en plug-in (gpsdecode[2]) konverteres dataene til *JavaScript Object Notification* (JSON) format. Til slutt lastes dataene inn i Kafka via prosessen *«PublishKafka_2_6»*.

[2] http://manpages.ubuntu.com/manpages/bionic/man1/gpsdecode.1.html



Figur 6: Eksempel på NiFi flow

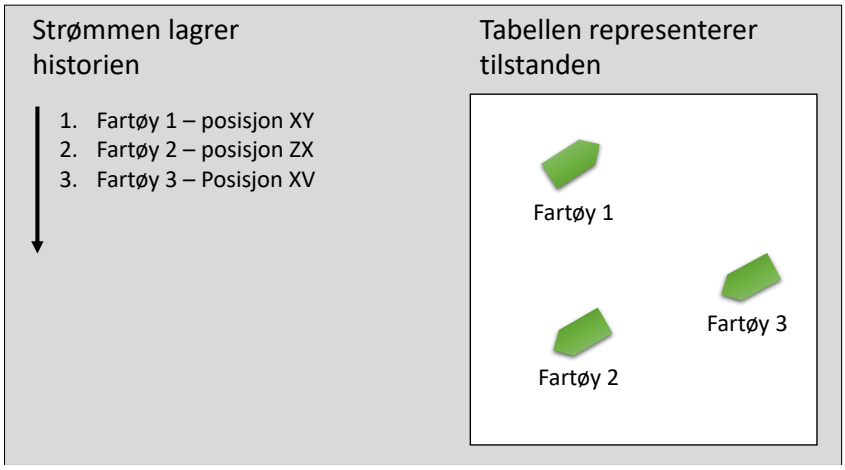
ksqlDB

ksqlDB er et verktøy for å gjøre sanntidsanalyser av strømmende data i Kafka meldingskøer (topics). ksqlDB er distribuert, skalerbart og prosesserer i sanntid (Narkhede, 2017). All datamanipulering og analyse i denne oppgaven kjøres i ksqlDB.

ksqlDB tilbyr et SQL-lignende spørrespråk for strømmende data i Kafka, noe som betyr at du kan skrive kontinuerlige spørringer som kjører på ubestemt tid. Det er to grunnleggende konsepter i ksqlDB som brukerne kan benytte når data behandles og analyseres. Avhengig av hvordan vi tolker meldingene i et Kafka *topic*, kan vi definere dem som enten strømmer eller tabeller.

Hvis vi ser på meldingene som kommer inn i et *topic* som en uavhengig sekvens av strukturerte data, tolker vi det som en strøm. Meldinger i en strøm har ikke noe forhold til hverandre og vil bli behandlet uavhengig. Hvis vi isteden betrakter meldingene som kommer inn i et *topic* som et sett med meldinger som oppdateres fortløpende, hvor en ny melding enten oppdaterer den forrige meldingen i settet med samme attributt, eller legger til en ny melding når det ikke er noen melding med tilsvarende attributt, så tolker vi temaet som en tabell (Jafarpour, Desai & Guy, 2019, s. 526).

For å eksemplifisere med dataene som er benyttet i eksperimentet: tabellen representer en tilstand på et gitt tidspunkt, så hvilke posisjoner fartøyene har på et gitt tidspunkt vil representeres av en tabell. Til sammenligning vil bevegelsene til fartøyene være en strøm med data (se figur 7). Mens dataene i tabellen vil endres fortløpende, vil ikke dataene i strømmen endres (Noll, 2020).

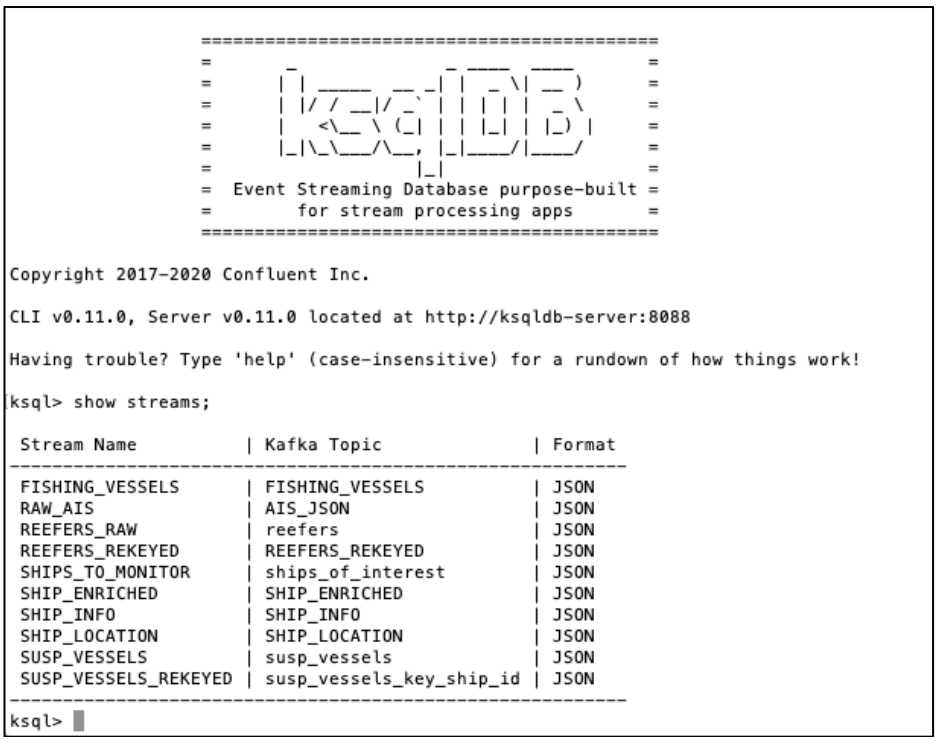


Figur 7: Forskjellen på strøm og tabell

ksqlDB tilbyr de fleste funksjonaliteter som et SQL har. Det vil si at man har mulighet til å manipulere data, gjøre statistiske beregninger, kombinere flere datasett og lese ut data (Noll, 2020). Den største fordelen med ksqlDB, slik jeg ser det, er at man ikke trenger å kunne kode for å bruke programvaren.

ksqlDB består av en serverkomponent og et kommandolinje-brukergrensesnitt. Gjennom brukergrensesnittet setter man opp spørringer som kontinuerlig kjøres mot den innkommende datastrømmen. Svarene fra spørringer skrives tilbake til Kafka som nye meldingskøer/topics.

Figur 8 viser et skjermutklipp fra ksqlDB og viser noen av de ulike strømmene som er satt opp for å gjennomføre analysene i dette eksperimentet.



Figur 8: Skjermutklipp av ksqlDB kommandolinje brukergrensesnittet

QGIS

QGIS er et geografisk informasjonssystem (GIS) som kan vise, redigere og analysere geografiske data og eksportere til de fleste vanlige kartfilformat. Programmet har verktøy for både vektor- og rasterkart. QGIS er gratis og åpen kildekode-programvare. QGIS visualiserer dataene og ble valgt fordi det er gratis, lett tilgjengelig og har funksjoner som gjør det enkelt å sortere og visualisere dataene.

Programvareversjoner

Følgende programvareversjoner ble benyttet i eksperimentet:

- Apache Kafka versjon 2.11-1.1.1
- QGIS versjon 3.17 Hannover
- ksqlDB versjon 0.14
- NiFi versjon 1.12.1

6.3 Hardware

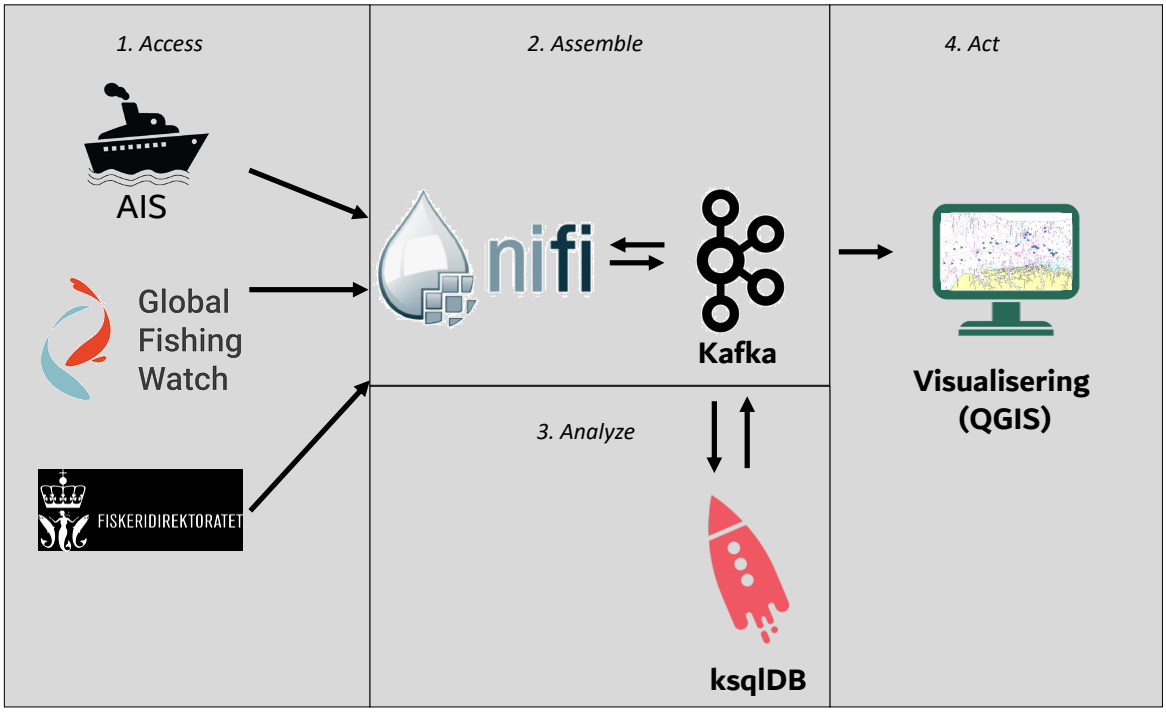
Maskinen som kjører stordatainfrastrukturen er en virtuell maskin med 32 GB RAM som kjører CentOS 7 som operativsystem. Maskinen kjører på en HP ProLiant DL380 Gen9 server med en 10-kjerners Intel Xenon E5-prosessor.

6.4 Oppsett og konfigurasjon

Programvaren kjører i såkalte Docker-containere. Docker gjør det mulig å pakke programvare med alle tilhørende tillegg og utvidelser i en «container» som kan flyttes rundt fra maskin til maskin uten behov for ytterligere oppsett. Docker-containere virker på tvers av plattformer, noe som gjør det mulig å kjøre Linux-programvare på f.eks. Windows eller Mac (Docker, 2021). Dette gjør det enklere å administrere og håndtere, da man ikke har behov for å installere og sette opp programvare og tjenester på tradisjonell måte. Klargjorte containere, såkalte *«images»* kan ofte fritt lastes ned fra mange ulike tilbydere på internett. Konfigurasjon av tjenestene og forholdet mellom dem blir definert gjennom en enkelt tekstfil, en såkalt *«docker-compose»*. Her konfigureres den enkelte container og hvordan nettverket dem imellom skal se ut. For å starte programmene kjøres en enkelt kommando, og hele infrastrukturen er etablert ila. minutter. Hvis en av containerne feiler eller det er behov for å gjøre endringer i oppsettet kan enkeltcontainere startes og stoppes individuelt.

6.5 Dataflyt i valgt stordatainfrastruktur

Enkeltkomponenter i seg selv utgjør ingen stordatainfrastruktur, det er først når det settes opp og kobles sammen, at vi kan hevde å ha etablert en stordatainfrastruktur. Figur 9 illustrerer hvordan dataene flyter, og hvilken rolle de ulike komponentene har. Som figuren viser har vi nå etablert en infrastruktur som ivaretar alle delprosessene i 4A, fra datakildene identifiseres til resultatet av analysene presenteres for sluttbruker.



Figur 9: Illustrasjon av dataflyten og softwarekomponenter

I steg 1 (*Access*) velges de ulike datakildene. For dette eksperimentet er det en kombinasjon av databaser fra Fiskeridirektoratet og Global Fishing Watch i comma-separated values (CSV) format og strømmende AIS-data fra Kystverket.

I steg 2 (*Assemble*) hentes disse datakildene inn i infrastrukturen ved hjelp av NiFi. I NiFi konverteres dataene og klargjøres for analyse ved at de blir konvertert til JSON-format før det skrives til Kafka som ulike topics. CSV er et filformat som bruker komma til å skille verdier (Wikipedia, 2020), mens JSON er en tekstbasert standard for å formatere dokumenter ifm datautveksling (Wikipedia, 2017). Kafka lagrer alle meldingene i meldingskøer, eller såkalte topics.

I steg 3 (*Analyze*) gjennomfører ksqlDB kontinuerlige spørringer og analyser mot topics i Kafka, for å bekrefte eller avkrefte de definerte indikatorene. De dataene som gir treff på de definerte indikatorene skrives deretter tilbake som nye topics i Kafka.

I steg 4 (*Act*) blir svarene fra analysene lest ut av noen definerte topics i Kafka. Av praktiske hensyn ble disse lest ut manuelt og deretter lagt inn i QGIS for visualisering. Dette muliggjør beslutninger og videre oppfølging.

7 Datasettet

Dette kapittelet gir grunnleggende informasjon om AIS som system og hvorfor dette er interessant som stordatakilde i eksperimentet som ble gjennomført. Den første delen av dette kapittelet vil omhandle AIS generelt, mens den siste delen tar for seg tidligere studier og påviste feilkilder med bruk av AIS som datagrunnlag.

AIS ble valgt som datakilde av flere grunner. For det første er dette data som er åpne og lett tilgjengelige (Kystverket, 2020b). For det andre tilfredsstiller AIS-data definisjonen av stordata, da dataene har høy hastighet og stort volum. I tillegg er dataene romlig-temporale, dvs. at de både har et geografisk og tidsmessig perspektiv. Dette gjør at dataene er anvendelige til mange typer analyser. For det tredje anses AIS som en viktig stordatakilde til bruk i det offentlige (Vivento & Kaupang, 2015, s. 33), og kan brukes fra alt til fiskerioppsyn til etteranalyse av ulykker til havs.

I sum gjør disse faktorene det svært interessant å bruke AIS som datakilde. Både for å bli kjent med AIS som datakilde, men også fordi dataene gjør seg svært godt som forskningsdata.

7.1 Automatic Identification System (AIS)

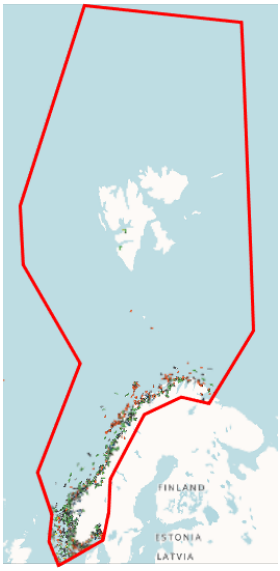
AIS ble innført av FNs sjøfartsorganisasjon International Maritime Organization (IMO) i 2004 som et antikollisjonssystem for å øke sikkerheten til sjøs. Systemet består av GPS-mottakere som logger skipets posisjon, og en VHF-sender som periodisk sender ut data og gjør dataene tilgjengelig (MarineTraffic, 2019, s. 7). AIS-mottakere langs kysten registrerer fartøy innenfor 40-60 nautiske mil, og med satellittbaserte systemer kan trafikk følges over alle hav.

Alle fartøy fra 300 bruttotonn og oppover i internasjonal fart, handelsfartøyer på 500 bruttotonn og oppover uavhengig av internasjonal fart, samt *alle* passasjerfartøy er pliktig å ha AIS installert ombord. (IMO, 2015, s. 1) I tillegg kreves det også på noen spesielle typer fartøy, eksempelvis må alle fiskebåter i EU over 15 m bære AIS (Eriksen, Greidanus & Delaney, 2018, s. 80).

I utgangspunktet ble systemet innført for å øke sikkerheten til sjøs, men har senere også fått stor betydning innen overvåkning av skipstrafikk – da basert på mottak av AIS-signalene på landstasjoner og via dedikerte AIS-satellitter. Norge har siden 2010 hatt operative AIS-satellitter i drift. I dag er det fire AIS-satellitter som brukes til å overvåke norske farvann (Kjerstad, 2020). Disse går i polarbane i ca. 600 kilometers høyde som gir dem lang rekkevidde og evne til å overvåke store havområder (Kleppe, 2015).

7.2 Tilgangen til datasettet

Kystverket tilbyr sitt AIS-nettverk *«AIS Norge»* i sanntid, enten som AIS rådata eller et nettbasert trafikkbilde i et kart. *«AIS Norge»* finnes i en åpen og en lukket del. Datatilgang gis i begge tilfeller over ordinær internettforbindelse. Denne studien benytter den åpne delen. Dette gir tilgang til AIS-data fra alle fartøy innenfor et dekningsområde som omfatter norsk økonomisk sone og vernesonene ved Svalbard og Jan Mayen, men med unntak av fiskefartøy under 15 meter og fritidsfartøy under 45 meter. Den røde streken i kartet under (Figur 10) viser den geografiske avgrensningen til datagrunnlaget.



Figur 10: Kartet viser den geografiske avgrensningen til datagrunnlaget

Dataene er gratis og tilgjengelig for alle, men er regulert under *Norsk Lisens for offentlige data* (NLOD, 2020). Det er ingen krav om brukerregistrering for å få tilgang. Dataene er tilgjengelig på IP adresse 153.44.253.27 port 5631 (Kystverket, 2020b).

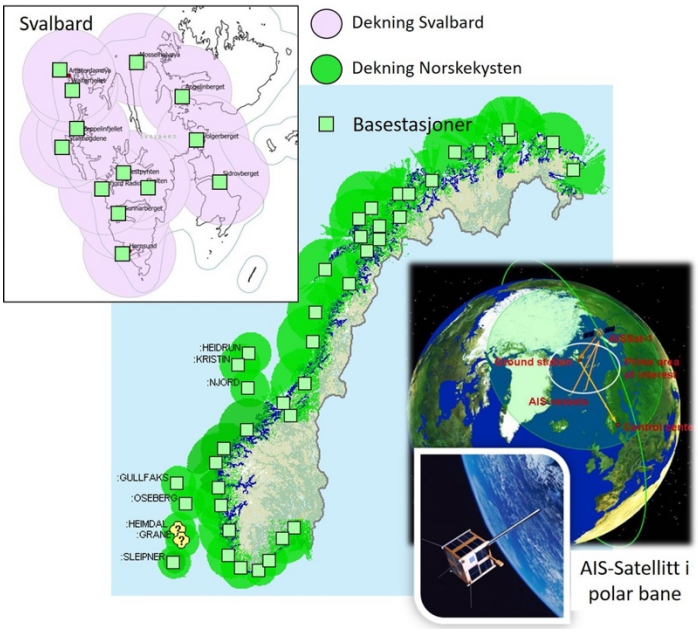
7.3 Tekniske karakteristikker

Systemet fungerer ved at AIS-transponderen ombord automatisk sender ut informasjon i fastsatte intervaller via en innebygget sender. AIS-data pakkes i standardiserte meldinger og sendes ut ved bruk av dedikerte kanaler i det maritime VHF båndet. Utsendingen skjer i et automatisk og selvorganiserende datasystem kalt *Self-Organized Time-Division Multiple Access* (SOTDMA) (Kjerstad, 2020).

Informasjonen stammer fra fartøyets navigasjonssensorer som f.eks. global satellittnavigasjon (GNSS/GPS) og gyrokompass. Dette mottas på fartøy i nærheten som igjen kan vise denne

informasjonen på dataskjermer og/eller kartplotter. Rekkevidden til AIS er begrenset av rekkevidden til VHF-båndet, noe som noe forenklet kan sies å begrenses av hvor høyt antennen er montert på skipet. Jo høyere montert, jo lenger rekkevidde. Typisk rekkevidde for et fartøy på sjøen er 20 nautiske mil (U.S. Coast Guard Navigation Center, 2016).

For å kunne benytte dataene til noe mer enn kollisjonsunngåelse til havs må dataene overføres fra fartøy til land. For å gjøre dette har Kystverket etablert et nettverk av landbaserte stasjoner som dekker, med enkelte unntak, området fra grunnlinjen og 40-60 nautiske mil ut fra kysten. Figur 11 viser omtrentlige dekningsområde. Disse basestasjonene sørger for å videresende AIS-data fra fartøy til Kystverket sentralt. Dette komplementeres av de nevnte AIS-satellittene, noe som medfører at dekningen i norske farvann er god (Johnsrud, Pedersen & Gravir, 2014, s. 30).



Figur 11: Dekningsområde for AIS-landstasjoner (kilde: SNL.no/AIS)

7.4 Meldingstyper:

Det er definert 27 forskjellige typer AIS-meldinger som kan sendes av AIS-transpondere. Generelt sender alle AIS-meldinger som et minimum ut følgende tre basiselementer:

- MMSI-nummeret
- Meldingstype
- En repeat-indikator

Data utover dette defineres av den enkelte meldingstype. Meldingstypene varierer fra fartøysinformasjon til værmelding og korreksjonssignaler for GNSS.

I denne oppgaven benyttes posisjonsrapporter som sendes fra fartøyet som meldingstype 1, 2, 3 og 27. Type 1, 2 og 3 viser at meldingen videresendes via bakkebaserte AIS-stasjoner, mens meldingstype 27 videresendes via satellitt. I tillegg benyttes meldingstype 5 som inneholder statiske data og data knyttet til seilasen.

Det faller utenfor oppgaven å forklare resterende meldingstyper, det henvises derfor til litteraturen for ytterligere forklaring av de resterende meldingstypene (U.S. Coast Guard Navigation Center, 2016).

7.5 Innholdet i AIS-meldinger

AIS-systemet foretar en forhåndsprogrammert utsending av to ulike former for data: statiske og dynamiske. Statiske data består av detaljer om fartøyet og om reisen og er definert som meldingstype 5. Informasjonen som sendes ut er Maritime Mobile Service Identity nummer (MMSI), kallesignal og navn, IMO-nummer, lengde og bredde, type fartøy, anløpshavn og lokalisasjon av posisjonsantenne (IMO, 2015, s. 5). Statiske data blir sendt ut automatisk hvert sjette minutt eller ved endring av data (Kystverket, 2020a).

Dynamiske data består av fartøyets posisjonsdata, fartøyets kurs «over grunnen», hastighet, retning, navigasjonsstatus (for eksempel *«til ankers»* eller *«underveis ved bruk av motorkraft»*), og *«rate of turn»* (antall grader kursendring per minutt). Dynamisk informasjon sendes med 3 minutter til 2 sekunders oppdateringsrate, avhengig av farten og kursendringene til fartøyet. Dynamiske data skal ikke kunne manipuleres av skipets besetning (Kystverket, 2020a).

Utdyping av sentrale datafelt

I tabellen nedenfor er en gjennomgang av de variablene som brukes videre i oppgaven.

Variabel	**Forklaring**
MMSI	Et fartøy identifiseres med navn og kallesignal. Et fartøys MMSI-nummer er en unik kode bestående av ni siffer som er internasjonalt entydig identifikasjonsnummer for hvert enkelt fartøy. Med bakgrunn i MMSI-nummeret er det mulig å avdekke fartøyets eier, og MMSI-nummeret kan endre seg ved eierskifte på et fartøy. (Schnelle, 2018, s. 44)

IMO	Et fartøys IMO-nummer er en unik kode bestående av syv siffer. I motsetning til MMSI- nummeret blir IMO-nummeret permanent tildelt fartøy uavhengig av eier. Regelverket omfatter alle fartøy på 100 bruttotonn eller mer, samt fiskefartøy, passasjerfartøy (ned til 12 meters lengde) og borerigger i internasjonal seilas. SOLAS-reguleringen fritar enkelte fartøy fra kravet. Dette gjelder fartøy som ikke har maskinell fremdrift, lystbåter, spesielle fartøy (eksempelvis SAR-fartøy), mudringslektere, flytedokker, krigsskip, og fartøy med konstruksjon av tre og som samtidig ikke er fiskefartøyer (IMO, 2017, s. 3). IMO-nummeret følger fartøyet gjennom dets levetid, og skal bestå uavhengig om enheten skifter eier. (Schnelle, 2018, s. 44)
SOG	Speed Over Ground er fart over grunnen og er den farten et fartøy beveger seg relativt i forhold til jorden. Farten beregnes ut ifra relative krefter som ytre påvirkning fra vind eller strøm. Hvis et fartøy har 13 knop gjennom vannet og motstrømmen er 3 knop vil fartøyet ha en fart på 10 knop over grunnen (SOG). (Schnelle, 2018, s. 44)
LAT & LON	Latitude og longitude er det samme som breddegrad og lengdegrad. Breddegrad brukes sammen med lengdegrad til å fastslå en presis stedsangivelse på Jordens overflate. (Hofstad, 2018)
Shiptype	Skipstype angis ved hjelp av en tosifret kode. Det første sifferet representerer den generelle kategorien til fartøyet, mens det andre sifferet, i noen tilfeller, representerer utfyllende informasjon angående type last. Første siffer: **1** = Reserved **2** = Wing In Ground **3** = Special Category **4** = High-Speed Craft **5** = Special Category **6** = Passenger **7** = Cargo **8** = Tanker **9** = Other

	Andre siffer: **0** = All ships of this type **1** = Major Hazard (Haz A) **2** = Hazard (Haz B) **3** = Minor Hazard (Haz C) **4** = Recognisable Hazard (Haz D) Eksempelvis betyr koden 30 at det er et fiskefartøy, mens 70 er lastefartøy. For en utfyllende liste anbefales hjemmesiden til MarineTraffic (MarineTraffic, 2018).

Tabell 1: Beskrivelse av sentrale variabler i AIS-meldinger som benyttes i oppgaven

7.6 Teknisk usikkerhet og svakheter i AIS-data

AIS-data er ansett for å være en pålitelig informasjonskilde og brukes som beslutningsgrunnlag verden over. Allikevel har systemet en del kritiske sårbarheter når man skal følge skipene. Systemet ble utviklet for å trygge navigering til sjøs og det ble derfor prioritert å lage en datastrøm som var lett tilgjengelig, og det ble derfor ikke fokusert på innebygde sikkerhetsmekanismer. Disse svakhetene blir i dag i økende grad utnyttet av fartøy for å skjule identitet, plassering og hvor de skal, enten for økonomisk vinning eller for å unngå nasjonale myndigheter (Kontopoulos, Chatzikokolakis, Zissis, Tserpes & Spiliopoulos, 2020, s. 87; MarineTraffic, 2019, s. 12; Windward, 2014, s. 2).

Selv om de fleste fartøy av en viss størrelse er pålagt å være utrustet med en AIS-sender, er man ikke pålagt å ha den påslått til enhver tid. Skipssjefen vurderer selv om fartøyet skal sende ut AIS-signal eller ikke. Dette gjør at fartøy kan «mørklegge» seg selv ved å skru av senderen, enten for å gjemme seg for myndighetene eller for å skjule for hvor man er. Sistnevnte kan gjøres av flere forskjellige grunner, for større tankere eller lasteskip kan man skru av senderen for å unngå piratangrep, mens for et fiskefartøy kan motivasjonen være å skjule gode fiskeplasser for andre fiskere (Kontopoulos et al., 2020, s. 86).

Presisjon og datakvalitet

Når man skal gjennomføre analyse på større datasett er det viktig å ha et bevisst forhold til kvaliteten på datagrunnlaget, altså hvilke feil og mangler finnes i datasettet. Dette er viktig fordi dette vil ha påvirkning på analyser og eventuelle algoritmer som analyserer dataene.

Tu et. al (2017) har gjennomført en studie som blant annet ser på datakvaliteten til AIS, og en av deres delkonklusjoner er at det er tre typer feil som man må være bevisst. For det første kan *«speed*

over ground»-verdiene ha urealistisk høye eller lave verdier, dette skjer pga støy i målingene eller i overføringen. Resultatet er verdier med ekstremt store positive eller negative verdier. Den andre typen er at samme AIS-melding repeteres. Dette skyldes at samme melding mottas av forskjellige AIS-basestasjoner, eller at fartøyet stanser opp, men fortsetter å sende ut AIS-meldinger kontinuerlig. Dette kan gi feil i plottingen av seilasen, og fartøyet kan få store hopp frem og tilbake da meldingene prosesseres fortløpende når de ankommer serveren. Den tredje er at AIS-meldinger forsvinner fordi protokollen har ikke har innebygde mekanismer for å verifisere at meldinger kommer fram. Samtidig er sendinger over VHF-båndet følsomme for støy fra andre elektriske innretninger som radarer (Tu et al., 2017, s. 1564).

En annen studie (Last, Bahlke, Hering-Bertram & Linsen, 2014) peker også på visse utfordringer i AIS-dataene, men det vurderes videre at dette ikke gjelder mer enn 0,25 % av utsendte data. Denne studien peker spesielt på utfordringene ved at noen av datafeltene som *«rate of turn»* og *«heading»*, ofte har feil verdier (Last et al., 2014, s. 794).

Kystverket har også analysert datakvaliteten til AIS ifm utarbeidelsen av *«Sjøsikkerhetsanalysen 2014».* Her pekes det på flere utfordringer som manglende dekning, brukerfeil og utstyrsfeil som kan oppstå. Allikevel vurderes disse faktorene å ha såpass liten betydning og innvirkning at disse feilkildene neglisjeres i deres analyse (Johnsrud et al., 2014, s. 30).

Manglende tidsstempel

Som nevnt tidligere, AIS er først og fremst et antikollisjonssystem og er designet for sanntidssynkronisering mellom fartøy til havs. Dette kan forklare hvorfor man ikke så det nødvendig å ha et komplett tidsstempel som en del av AIS-meldingsformatet. Ved å utelate denne informasjonen reduseres overføringsmengden og overføringstiden. Grunnet denne mangelen på tidsstempel må dette legges til når AIS-meldingen ankommer datasystemet som skal analysere dataene. Dette gjør at tidsstempelet man jobber med representerer når man mottok dataene, ikke når de ble sendt ut fra fartøyet. Dette kan i verste fall føre til at meldingen ankommer mottakeren flere timer etter at det ble sendt fra fartøyet (Skauen, 2016, s. 530). Dette medfører at presisjonen til tidsstempelet må tas høyde for når analyser og algoritmer skal designes.

Årstidsvariasjoner

Norske farvann har til enhver tid en betydelig fiskeaktivitet, men fisket varierer med årstid. Fangsten fra kystflåten er i stor grad avhengig av tilgjengeligheten på fisken, og har dermed store sesongvariasjoner. I januar og februar fisker trålerne etter torsk og hyse utenfor kysten av Finnmark.

Noen fartøy går ned mot kysten av Lofotenområdet etter hvert som torsken kommer inn mot land (Oksnes, 2014, s. 16).

Fisket blir også påvirket av vær og vind. Som et eksempel traff ekstremværet «Frank» norskekysten rett før innsamlingsperioden (20.-22. januar 2021) til eksperimentet skulle starte (Meteorologisk institutt, 2021). I denne tidsperioden var svært få fiskefartøy aktive i norske farvann og de fleste fartøy lå til kai og ventet på bedring av været. Slike variasjoner kan gi utslag på analysene og er noe man må være bevisst når algoritmer utformes og svarene tolkes.

7.7 Manipulasjon av AIS-data

En studie gjennomført av det israelske selskapet Windward (2014), et firma som leverer etterretningsløsninger basert på AIS-data, har laget en oversikt over hvordan man kan manipulere AIS-data. Funnene fra deres rapport er gjengitt nedenfor.

Falsk identitet

Identiteten til et fartøy angis med informasjonen (MMSI, IMO nummer) som er beskrevet tidligere i dette kapittelet. Studien viser at bruken av falsk eller stjålet identitet er et økende problem, og at 1% av fartøyene på verdensbasis er berørt (Windward, 2014, s. 4).

Skjuler destinasjon

Studien hevder at fartøyer bare i 41% av tiden rapporterer endelig destinasjon. Denne informasjonen skal være inkludert i data som AIS-transponderen sender ut (Windward, 2014, s. 4).

Slår av AIS – «going dark»

Rapportens funn viser at mer enn 25% av fartøyene skrur av AIS i minst 10% av tiden. Dette er mer vanlig blant fartøy på over 250 meter. Windward peker på at dette indikerer at fartøyene som bærer den største lasten har større insentiver til å skjule sine aktiviteter på enkelte tidspunkt (Windward, 2014, s. 4). En analyse fra samme firma i 2020 viser at antallet fartøy som slår av AIS nå er nedadgående, spesielt i området rundt Iran. Dette er på grunn av økt fokus på denne «mørkleggingen» og at dette nå alene kan være grunn til å havne på sanksjonslister (Primor, 2020).

Manipulering av GPS data

AIS-senderen foretar ingen validering av GPS-data. Derfor vil en hvilken som helst posisjon som blir gitt inn i AIS-senderen bli sendt ut som skipets posisjon, uavhengig av skipets faktiske posisjon. Denne type manipulering kan få det til å se ut som et fartøy er på helt andre steder enn der det virkelig befinner seg (Windward, 2014, s. 5).

Forfalske AIS-signalet («spoofing»)

Ifølge rapporten har det blitt påvist at en kan hacke AIS-datastrømmen og eksempelvis generere «spøkelsesskip» – fartøy som i virkeligheten ikke eksisterer. Videre kan man ved å *«spoofe»* signaler generere falsk støy i AIS-bildet som negativt kan påvirke den maritime situasjonsforståelsen (Windward, 2014, s. 5).

7.8 Oppsummering

AIS er en god datakilde som passer godt for å analysere problemstillinger knyttet til det maritime domenet, men allikevel er ikke dataene uten utfordringer. For eksperimentet som ble gjennomført i denne oppgaven er det spesielt utfordringene knyttet til feil med *«speed over ground»* og at samme melding leses flere ganger som kan gi unøyaktigheter. I tillegg viser funnene fra Windward-rapporten at det finnes muligheter for bevisste og ubevisste menneskelige feil. Disse feilkildene er med på å påvirke datagrunnlaget og analysene i denne oppgaven. Dette er viktig å ta med seg videre i arbeidet med tolkning og analyse av slike data. I tillegg er det viktig at algoritmer designes på en slik måte at de er robuste nok til å håndtere feilkildene i datagrunnlaget.

8 Eksperimentet

I dette kapittelet vil eksperimentdesignet bli presentert. Eksperimentets funksjon er å teste, undersøke og prøve ut om nyskapningen tilfredsstiller artefaktbehovet (Stølen, 2019, s. 152). Artefaktbehovet, som ble presentert i kapittel 2, er: «*Utvikle et verktøy som i sanntid analyserer en kontinuerlig stordatastrøm, og fortløpende sjekker om disse dataene tilfredsstiller et sett med forhåndsdefinerte indikatorer og hypoteser, og svarer dette til bruker fortløpende».* Verktøyet som skal evalueres består av to komponenter. Den første komponenten er metoden som ble presentert i kapittel 5, stordatadrevet ACH. Den andre komponenten er stordatainfrastrukturen som ble presentert i kapittel 6. Eksperimentet vil benytte seg av AIS som stordatakilde med de fordeler og ulemper det har, slik det ble redegjort for i kapittel 7.

Innledningsvis vil rammene for eksperimentet presenteres, herunder bakgrunn for problemstilling og selve problemstillingen. Videre benyttes stordatadrevet ACH som metode og ved hjelp av denne vil problemstillingen i eksperimentet brytes ned til indikatorer. Disse brytes videre ned til algoritmer slik at stordatainfrastrukturen kan brukes til å svare på problemstillingen.

8.1 Problemsett

For å teste metodikken er det behov for et problemsett[3] som kan besvares ved å benytte AIS som datakilde. I en *Norsk Offentlig Utredning* (NOU) fra 2019, om framtidens fiskerikontroll, omtales det flere utfordringer med dagens fiskeriforvaltning. Her pekes det blant annet på utfordringer med ulovlig, urapportert og uregulert (UUU) fiske. UUU-fiske er en samlebetegnelse på delvis overlappende fiskepraksis som skjer i strid med fiskerilovgivning og rapporteringsplikt, samt fiske hvor slike bestemmelser mangler (Pedersen et al., 2019, s. 76-78).

UUU-fiske har fått betydelig oppmerksomhet internasjonalt de senere årene, fordi omfanget av UUU-aktiviteter på global basis er en stor trussel mot fiskebestander, marine økosystemer, forsvarlig fiskeriforvaltning, kystsamfunn og konsumentenes tiltro til fisk som et etisk produkt (Nærings- og fiskeridepartementet, 2018).

En studie fra den amerikanske tenketanken *«Center for Advanced Defense Studies»* (C4ADS) har analysert problematikken runde UUU-fiske i stor detaljgrad (Brush, 2019). De hevder at deres funn er

[3] Uttrykket problemsett brukes bevisst for å skille mellom oppgavens problemstilling og problemstillingen som benyttes i eksperimentet.

gyldige globalt, et syn som også støttes av Kroodsma, Miller & Roan (2017). Problematikken rundt UUU-fiske anslås å være størst i fattigere land med fiskeritradisjoner og -interesser.

I norske farvann hevder Pedersen et.al (2019, s. 76-78) at UUU-fiske ikke er en stor utfordring i dag, men likevel knyttes det endel usikkerhet til utenlandske fiskefartøy generelt og omlastingsaktivitet til havs spesielt. Samtidig er det ikke lenge siden man hadde store utfordringer også i norske farvann. I årene 2002–2005 viste beregninger fra Fiskeridirektoratet at det urapportert i snitt ble tatt opp 100 000 tonn norsk-arktisk torsk i Barentshavet hvert år (Nærings- og fiskeridepartementet, 2018).

Da det viser seg å være noe usikkerhet angående omfanget av UUU-fiske i norske farvann synes dette å være et godt utgangspunkt. Denne studien ønsker derfor å analysere denne problemstillingen ved å benytte stordatadrevet ACH som beskrevet tidligere. Følgende problemsett nyttes derfor videre i eksperimentet:

Foregår det ulovlig, urapportert og uregulert (UUU) fiske i norske farvann?

8.2 Praktisk bruk av stordatadrevet ACH

Steg 1: Utvikle hypoteser

Hypotesene skal spenne ut og representere handlingsrommet. Grunnet problemsettets binære natur er det ikke behov for å operere med fler enn to hypoteser. Hypotesene bekrefter eller avkrefter hvorvidt det foregår UUU-fiske i norske farvann.

H_0: Nei, det foregår ikke UUU-fiske i norske farvann

H_1: Ja, det foregår UUU-fiske i norske farvann

Steg 2: List opp signifikante bevis

Nevnte C4ADS-studie har identifisert seks observerbare indikatorer som i sum øker sannsynligheten for at fiskefartøy driver UUU-fiske (Brush, 2019, s. 15). En indikator alene er altså ikke nok til å falsifisere hypotesene, men treff på flere indikatorer på flere fiskefartøy vil til sammen øke sannsynligheten for at det foregår UUU-fiske i norske farvann.

I tabellen under er indikatorene listet opp.

Indikator	Beskrivelse
Historikk med UUU-fiske	Fartøy eller fartøyeier er involvert i tidligere UUU-aktivitet
Omlasting til sjøs	Møte mellom fartøy, spesielt mellom fiskefartøy og transportfartøy med kjølemulighet, for å overføre fangst, forsyninger og annet.
«Mørk» AIS-aktivitet	Manipulering av posisjonsdata eller at AIS slås av
Flaggmanipulering	Seile med såkalt bekvemmelighetsflagg, hyppig omflagging til andre stater, eller seile uten flagg i det hele tatt
Bekvemmelighetshavn	Bruken av private eller offentlige havner med lite eller ingen tilsyn
Endre fartøysidentifikasjon	Fartøy som gjentakende endrer fartøysnavn, kallesignal eller fysisk utseende

Tabell 2: Indikatorliste fra C4ADSs rapport *«Strings Attached»* (Brush 2019, s. 15)

Alle indikatorene nevnt ovenfor er gode indikatorer, men grunnet tilgjengelig tid og for å redusere studiens kompleksitet avgrenses eksperimentet til å svare på tre av de seks indikatorene ovenfor. Dette er historikk med UUU-fiske, omlasting til sjøs og «mørk» AIS-aktivitet. Disse tre velges da de vurderes å ha størst potensiale for å kunne besvares ved hjelp av analysemetodene og datasettet lagt til grunn. Før videre arbeid tydeliggjøres det hva hver enkelt indikator skal svare ut (se tabell 3).

Indikator	Forklaring
Historikk med UUU-fiske	Fiskefartøy som har en historikk med ulovlig, urapportert og/eller uregulert fiske øker sannsynlighet for at fartøyet foretar med UUU-fiske.
Omlasting til sjøs	Fiskefartøy som møter transportfartøy (med kjølemulighet) for å overføre fangst eller tilsvarende utenfor havner, øker sannsynligheten at fartøy bedriver UUU-fiske.
«Mørk» AIS-aktivitet	Fiskefartøy som skrur av AIS i en lenger tidsperiode øker sannsynligheten for at UUU-fiske foregår.

Tabell 3: Reformulerte indikatorer.

Steg 3: Vurder diagnostisk score

For å vurdere hvor godt de ulike indikatorene avkrefter eller bekrefter de ulike hypotesene vurderer vi den diagnostiske scoren. En indikator med høy diagnostisk score vil avkrefte/bekrefte kun én hypotese, mens en indikator med lav diagnostisk score vil bekrefte/avkrefte flere.

Indikator	H_0	H_1
Historikk med UUU-fiske	Inkonsistent	Konsistent
Omlasting til sjøs	Inkonsistent	Konsistent
«Mørk» AIS-aktivitet	Inkonsistent	Konsistent

Tabell 4: Vurdering av den diagnostiske scoren til den enkelte indikator

Som tabellen over viser vurderes alle indikatorene som gode nok til å benyttes i det videre analysearbeidet da alle falsifiserer H_0.

Steg 4: Revurder hypotesene

Med bakgrunn i analysen av indikatorene, samt gjennomgang av hva datasettet kan forvente å gi oss av data, kommer det fram at både data og de nevnte indikatorene er på fartøynivå. For å få svar på problemsettet er man derfor avhengig av å se på det enkelte fartøy, med andre ord se på hvilke spesifikke fiskefartøy som har økt sannsynlighet for UUU-fiske. For å gjøre rede for denne nyansen må hypotesene endres noe. De opprinnelige hypotesene må derfor reformuleres for å fange opp denne endringen. De reformulerte hypotesene blir derfor:

H_0: Nei, det finnes ikke fiskefartøy som driver UUU-fiske i norske farvann

H_1: Ja, det finnes fiskefartøy som driver UUU-fiske i norske farvann

Dette betyr at de tre indikatorene må testes mot alle fiskefartøyene som er tilgjengelig i datastrømmen. For å svare på problemsettet, foregår det UUU-fiske i norske farvann, trengs det i prinsippet kun å få bekreftet at ett fartøy driver med UUU-fiske i norske farvann.

Steg 5: Velg stordatainfrastruktur

Dette delsteget dekkes i sin helhet i kapittel 6 – Stordatainfrastruktur.

Steg 6: Bryt ned indikatorer

For å besvare indikatorene må de konkretiseres i enda større grad, og brytes ned slik at programvaren, i dette tilfellet ksqlDB, kan lete etter indikatoren og bekrefte eller avkrefte den.

Indikator 1: Historikk med UUU-fiske

I forbindelse med det globale fokuset på å bekjempe ulovlig fiske har man blitt enige om en rekke internasjonale forvaltnings- og kontrolltiltak. Blant disse finner man såkalte «svartelister» som gir oversikt over fartøy som tidligere har vært involvert i UUU-fiske (Pedersen et al., 2019, s. 31).

Nasjonalt er det Fiskeridirektoratet som forvalter den norske svartelisten. Å bli oppført på svartelisten medfører at irreversible sanksjoner kan ilegges. Utenlandske fartøy som havner på den norske svartelisten, kan bli nektet lisens i norsk økonomisk sone. Listen er åpent tilgjengelig på Fiskeridirektoratets nettsider (Fiskeridirektoratet, 2020).

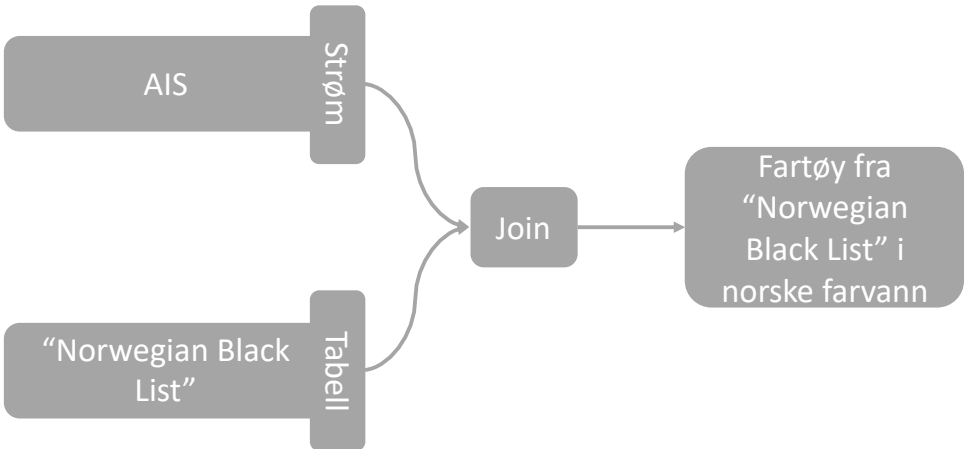
IMO number	Current name	Previous name(s)	Current flag	Previous flag(s)	Current call sign	Previous call sign(s)	Date included in the list - Norway	Comments/ Information confirmed in Lloyd's Register
	Ocean		Angola		D3R21		27/11/1998	
	Rex/Arnar 2		Cyprus		P3XV4		27/11/1998	
	Solea		Estonia				27/11/1998	
IMO 7324376	Arnarborg	Polmar Eht, Arnarborg, Hagangur 1	Latvia	Iceland, Latvia, Belize	YLDV	V3MT5	27/11/1998	Stern trawler. Ok Lloyd's
IMO 6812986	Senator	Otto, Prestland, Dalborg, Ottar Birting	Latvia	Faeroe Islands, Iceland, Panama	YLAC		27/11/1998	Stern trawler. Ok Lloyd's
IMO 5403740	Dux	Anuva, Albri, High Sierra	Lithuania	Sierra Leone, Belize	LYMP	9LFHS, V3NHV	27/11/1998	Fishing vessel. Ok Lloyd's

Tabell 5: Utdrag fra Fiskeridirektoratets svarteliste

Listen (se tabell 5) har flere ulike attributter som IMO-nummer, navn, tidligere navn, flagg, tidligere flagg, kallesignal, tidligere kallesignal, dato for når fartøyet havnet på listen, samt kommentarer.

For å besvare indikatoren vil fartøyene på den norske svartelisten kontinuerlig sjekkes mot AIS-datastrømmen for å finne ut om noen av disse er i norske farvann. Innledningsvis vil derfor listen lastes ned fra Fiskeridirektoratet sine nettsider, konverteres fra PDF til CSV-format, og lastes opp i en egen meldingskø i Kafka. Det settes deretter opp en spørring i ksqlDB som kontinuerlig sjekker om den finner noen av IMO-numrene i «svartelisten» i den løpende AIS-strømmen (se figur 12). Dette gjøres kun mot IMO-nummeret grunnet tekniske begrensninger i programvaren som p.t. kun gjør det mulig å kjøre slike operasjoner mot kolonner med numeriske verdier. Ideelt sett burde man kjørt en sjekk mot flere av feltene som navn, kallesignal etc. Det er verdt å merke seg at kun 78 av 112 fartøy på listen har IMO-nummer. Dette er ikke optimalt da det er 34 fartøy som man ikke kan identifisere med denne metoden. Allikevel vurderes det å være godt nok for dette eksperimentet, da ett treff alene vil være godt nok til å verifisere metoden.

Ved treff på indikatoren vil disse fartøyene bli skrevet til en ny tabell som kontinuerlige oppdateres med siste rapporterte posisjon på fartøyet. Denne tabellen kan man gjøre oppslag mot når man ønsker status på indikatoren.



Figur 12: Visualisering av analysemetodikken for å finne fartøy i norske farvann med UUU-historikk

Indikator 2: Omlasting til sjøs

Omlasting til sjøs defineres i denne konteksten til å være når et transportskip mottar fisk fra et fiskefartøy et annet sted enn i havn. Omlasting er både en logistisk og økonomisk nødvendig aktivitet i mange fiskerier. Samtidig er det verdt å merke at uten effektive kontrolltiltak, utgjør omlasting en alvorlig trussel mot fellesskapets ressurser. Omfattende omlastingsaktivitet var bl.a. med på å fasilitere det betydelige russiske overfisket som fant sted rundt årtusenskiftet i norske og internasjonale farvann, der enorme mengder ulovlig og urapportert fisk ble tatt (Pedersen et al., 2019, s. 77).

Global Fishing Watch (GFW), en uavhengig, internasjonal ideell organisasjon, som jobber med å kartlegge kommersielt fiske på et globalt nivå, har også fokusert på problematikken med omlastingsaktivitet. Deres analyser antyder også at det sannsynligvis foregår endel omlastingsaktivitet i norske farvann (Kroodsma et al., 2017, s. 9).

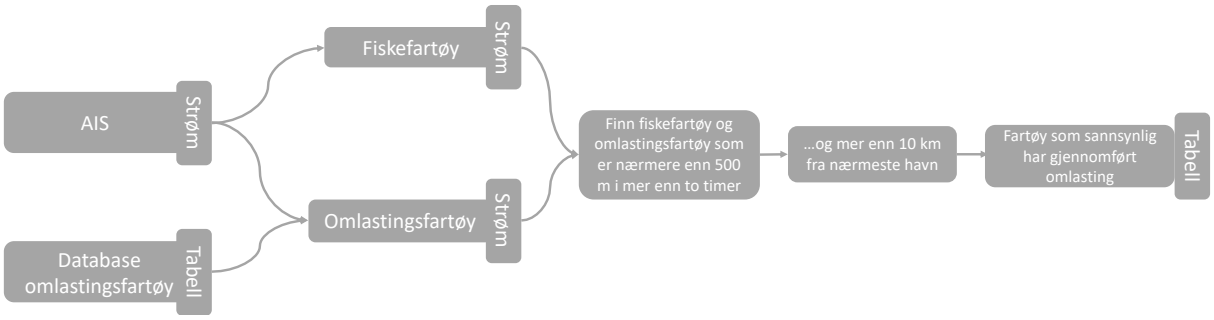
Analysene til GFW tar utgangspunkt i omlastingsfartøyene og sporer disse ved hjelp av AIS. Omlastingsfartøyene er samlet i en egen database som kan lastes ned fra GFW sin hjemmeside[4]. Denne databasen er benyttet i eksperimentet.

En sannsynlig omlasting defineres ut fra følgende kriterier: hvis et fiskefartøy og et omlastingsfartøy er nærmere hverandre enn 500 meter i minst to timer og har en fart under to knop, og dette er minst 10 km fra nærmeste havn. Disse seleksjonskriteriene balanserer behovet for å detektere fartøy som er nær hverandre i lengre perioder samtidig som man tar høyde for utfordringer med AIS-dekning og

[4] https://globalfishingwatch.org/data-download/datasets/carriers:v20201201

evt. utfordringer med manglende satellittdekning. Vi utelukker også fartøy som ligger ved siden av hverandre i havn. (Kroodsma et al., 2017, s. 7).

For å besvare denne indikatoren gjøres følgende (se figur 13). Innledningsvis lastes databasen ned fra GFW sine hjemmesider, konverteres og lastes opp i Kafka som en topic. Deretter etableres det en datastrøm for fiskefartøy, og en for omlastingsfartøy. Videre kjøres disse strømmene mot hverandre for å filtrere ut de fartøyene som er nærmere enn 500 m og har fart under to knop. Så sjekkes det hvor lenge fartøyene har ligget ved siden av hverandre og de som har ligget nærmere hverandre enn 500 meter i mer enn to timer skrives til en ny topic. Avslutningsvis sjekkes det at fartøyene er minst 10 km fra nærmeste havn.



Figur 13: Visualisering av analysemetodikk for å identifisere mulig omlastingsaktivitet

Indikator 3: «Mørk» AIS-aktivitet

Å skru av AIS for å skjule lyssky aktivitet er en enkel og effektiv metode for å unngå sporing. Det er som tidligere nevnt ikke ulovlig å skru av transponderen, og skipssjefen har anledning til å bestemme om man skal seile med den av eller på. I tillegg er det ikke 100 % AIS-dekning i norske farvann. Grunnene til at man ikke rapporterer via AIS kan derfor være mange. Det å si med sikkerhet om transponderen skrus av for å skjule UUU-fiske, om dekningen er dårlig eller om skipssjefen skrur av for å skjule gode fiskeplasser for andre fiskefartøy i nærheten er derfor umulig å si.

Det er gjort mye forskning på å finne metoder for å detektere når AIS skrus av med hensikt kontra når fartøyene kommer utenfor AIS-dekning. Kontopulos et al. (2020) har foreslått et rammeverk for sanntidsdeteksjon som kan skille mellom når fartøy har dårlig AIS-dekning og når AIS skrus av med hensikt. Ved å implementere et slikt rammeverk vil man med et høyere presisjonsnivå kunne si noe om hvorvidt fartøyet har skrudd av AIS eller ikke, men man sier fortsatt ikke noe om hensikten. Utfordringen med denne indikatoren er derfor å skille mellom når man skrur av AIS med gode og dårlige hensikter.

For det videre arbeidet med å besvare denne indikatoren forutsettes det at er god AIS-dekning i norske farvann (Johnsrud et al., 2014, s. 33). Gitt denne forutsetningen kan man derfor vurdere at i de tilfellene hvor det er utfall av AIS i lengre tidsperioder, så er AIS skrudd av med vilje. Denne metoden vil fange alle AIS-utfall og diskriminerer sånn sett ikke mellom gode og onde hensikter, men vil være godt nok for å teste metoden.

En intendert AIS-mørklegging defineres derfor hvis et fartøy ikke har rapportert posisjon via AIS siste seks timer. Det vil med denne metoden tas høyde for utfordringer med dekning for bakkestasjoner og evt. utfordringer med satellittdekning.

For å besvare denne indikatoren (se figur 14) etableres det en tabell i ksqlDB som oppdateres fortløpende med posisjon og tidspunkt for hvert enkelt fiskefartøy. Det gjøres deretter spørringer mot denne tabellen og fartøy som ikke har rapportert posisjon på AIS siste seks timer vil listes opp.

Figur 14: Visualisering av analysene for å identifisere mørklegging av AIS

Steg 7: Hent inn data

Når stordatainfrastrukturen er etablert og algoritmene er satt opp i analyseverktøyet starter selve innhentingsoperasjonen. I innhentingsfasen vil programvaren søke etter svar på de ulike indikatorene i den løpende AIS-strømmen. For eksperimentet i denne oppgaven ble det gjennomført to innhentingsfaser som varte i hhv 18 og 36 timer.

Steg 8: Presenter funnene

Datamengdene i stordataløsninger kan være betydelige. Eksempelvis produserer løsningen som er benyttet i denne studien ca. 5 millioner meldinger i døgnet. For at analysene som gjøres skal ha en effekt er det viktig av informasjonen presenteres for analytikere på en oversiktlig og relativt selvforklarende måte.

Da løsningen er designet for å analysere dataene i sanntid vil treff på indikatorene fortløpende publiseres i en rekke topics i Kafka. Informasjonen fra analysene hentes ut fra Kafka og presenteres ved hjelp av tabeller og kart. Det presenteres en oversikt over hvor fartøyet er, samt at analysegrunnlaget vises. I dette eksperimentet krever denne fasen noe manuelt arbeid og tilrettelegging av dataene for å flytte dem fra Kafka til hhv. QGIS og Excel for å lage hhv. kart og tabeller.

Steg 9: Formidle

Siste steg er å vurdere den rapporterte informasjon og avgjøre om informasjonen er av en slik karakter at den er av interesse for beslutningstakerne. For dette eksperimentet blir funnene presentert i kapittel 9.

I en annen kontekst vil det være naturlig å presentere funnene for en beslutningstaker eller kolleger slik at man får besluttet hva som skal gjøres på bakgrunn av funnene. I dette eksperimentet som ga treff på ett bestemt fiskefartøy kan man eksempelvis sende Kystvakta for å inspisere fartøyet, eller beslutte at man setter spesielt fokus på dette fartøyet en periode fremover.

9 Resultater

I dette kapittelet vil resultatene fra gjennomføringen av eksperimentet bli presentert. Det vil innledningsvis bli redegjort for datagrunnlaget i eksperimentet før resultatene fra analysene blir presentert. Et representativt utvalg av resultater fra analysene vil bli presentert og visualisert.

Datagrunnlaget er fra to tidsperioder. Den første tidsperioden er fra kl. 08:00 27. januar til ca. kl. 22:00 29. januar. Dette datagrunnlaget brukes til å svare på «Indikator 1 - Historikk med UUU-fiske» og «Indikator 2 - Omlasting til sjøs». Den andre tidsperioden er fra 5. februar kl. 18:00 til 6. februar kl. 07:00 og brukes til å svare på «Indikator 3 - «Mørk» AIS-aktivitet».

9.1 Funn indikator 1: Historikk med UUU-fiske

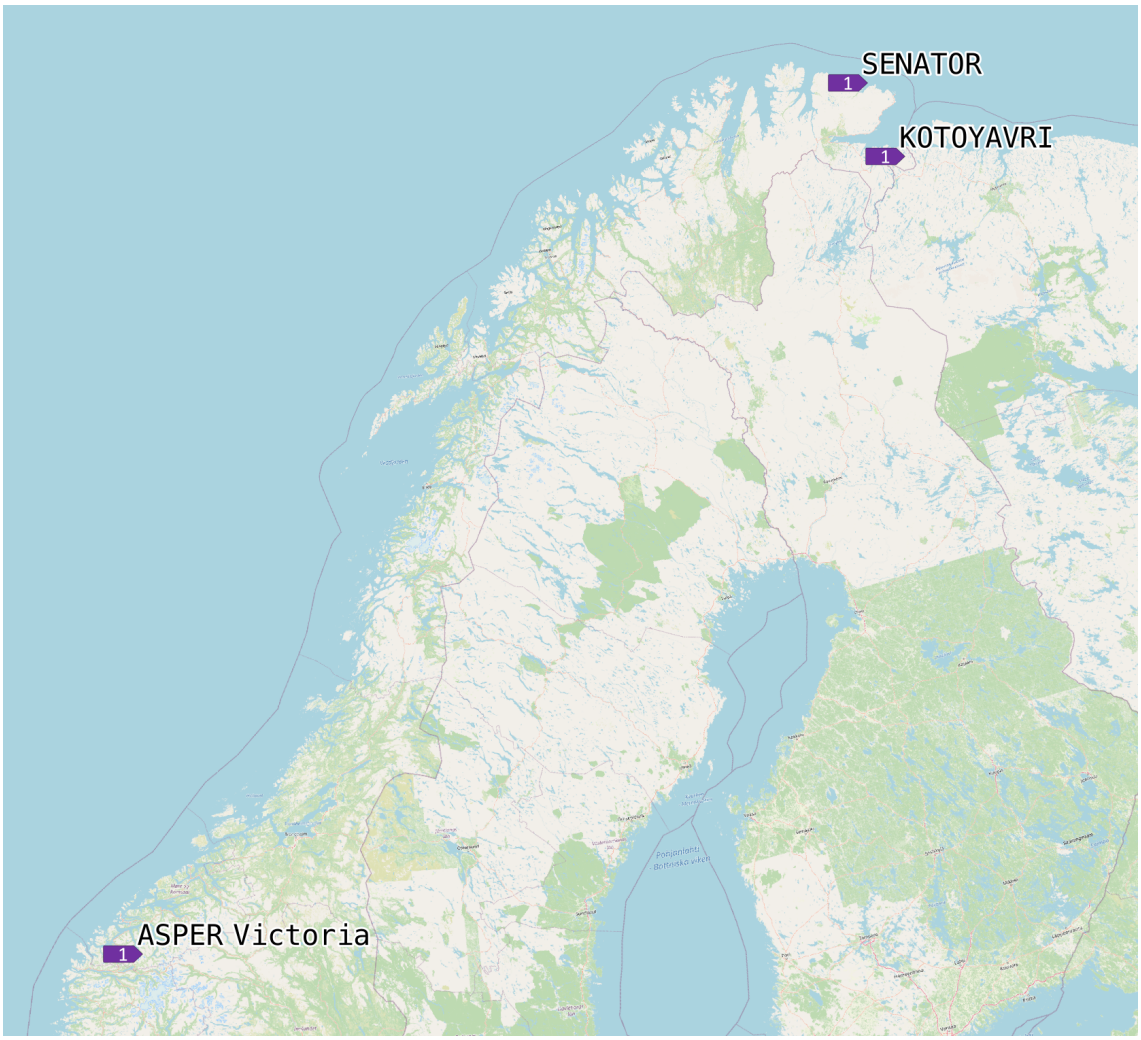
Indikator 1 er gyldig hvis et fartøy i norske farvann også er registrert på den norske svartelisten. I løpet av innhentingsfasen var det totalt tre fiskefartøy som sendte ut AIS-signal og samtidig var på svartelisten. Dette var fartøyene:

1. Asper Victoria
2. Senator
3. Kotoyavri

Kartet og tabellen under viser disse tre fartøyene. Aktiviteten til disse fartøyene deles i to kategorier, ingen aktivitet og sannsynlig fiskeriaktivitet. Fartøyene, «Asper Victoria» og «Senator», har ligget til ro ved kai i rapporteringsperioden så dette kategoriseres som ingen aktivitet. Aktiviteten til «Kotoyavri» kategoriseres som sannsynlig fiskeriaktivitet da dette fartøyet har seilt inn til Kirkenes, ligget til ro ved kai, før fartøyet har seilt til havs igjen.

IMO	MMSI	SHIPNAME	SHIPTYPE	LAT	LON	SPEED	TIMESTAMP	Treff på indikator
5225851	257390000	ASPER Victoria	Other	62.0747	6.21417	0.0	27-01-2021 08:48:38	Ja
6812986	275171000	SENATOR	Fishing	70.6374	29.7318	0.0	27-01-2021 08:48:01	Ja
7002368	273440730	KOTOYAVRI	Fishing	69.7304	30.0751	4.5	27-01-2021 08:48:44	Ja

Tabell 6: Tabellen over viser funn fra indikator 1 27. januar kl. 08:48



Figur 15: Kartet viser hvor de tre fartøyene som er på svartelisten var lokalisert 27. januar kl. 08:48

9.2 Funn indikator 2: Omlasting til sjøs

Funn på indikator 2 er definert som hvis et fiskefartøy og et omlastingsfartøy er nærmere enn 500 m og hastigheten er under 2 knop i mer enn to timer, og de samtidig er mer enn 10 km fra nærmeste havn.

Det tekniske aspektet for å få uttelling på denne indikatoren er betydelig mer komplekst enn de to andre indikatorene. Grunnet noe begrenset støtte for geografiske analyser i programvaren måtte analysen derfor gjøres i tre delanalyser. Første delanalyse er å identifisere kombinasjoner av fiskefartøy og omlastingsfartøy som er nærmere enn 500 m og som holder en fart under 2 knop. I andre delanalyse vil alle fartøyskombinasjoner som ivaretar dette kriteriet, deretter analyseres om de gjør dette i lenger enn to timer. De fartøyskombinasjonene som gjør dette skrives deretter til et nytt topic. Del tre er å identifisere hvilke av fartøyskombinasjonene som er lenger enn 10 km fra

nærmeste havn. Dette ble gjort manuelt ved å hente ut resultatene fra delanalyse 2 fra Kafka, legge dette inn QGIS og deretter måle avstanden fra fartøyskombinasjonene til nærmeste havn.

Det ble hentet ut totalt fem fartøyskombinasjoner (se figur 16) som resultat av delanalyse 1 og 2, men etter delanalyse 3 var det kun én kombinasjon som oppfylte alle tre kriterier i algoritmen. Tabell 7 viser de fem fartøyskombinasjonene som var resultat av delanalyse 1 og 2, men som vist i feltet «Avstand havn» er fire av funnene lokalisert i havner og er derfor ikke gyldige.

Figur 16: Kartet viser resultatet fra delanalyse 1 og 2

Den eneste fartøyskombinasjonen som gir utslag på indikatoren er et møte mellom fiskefartøyet Mys Slepikovskogo og omlastingsfartøyet Belomoyre i Recherfjorden, en fjordarm på Svalbard. Selve stedet for den mulige omlastingen ligger ca. 80 km fra Longyearbyen.

Fiskefartøy	FF MMSI	Omlastingsfartøy	OF MMSI	LAT	LON	Timer saml	Avstand havn	Treff på indikator
MYS SLEPIKOVSKOGO	273897000	BELOMORYE	273148810	77.5128	14.6267	4,8	81 km	Ja
EIGENES	258216000	HORISONT III	257349500	58.0769	8.0138	4,8	0 km	Nei
BRIS	257608500	HORISONT III	257349500	58.077	8.013	4,7	0 km	Nei
NEREY	273433400	KATRAN	273440860	69.7274	30.0314	4,5	0 km	Nei
KINGS BAY	258654000	ROFJORD	57430000	62.3237	5.6635	4,5	0 km	Nei

Tabell 7: Resultat fra indikator 2, merk at det kun er fartøyet Mys Slepikovskogo som oppfyller alle kriterier og gir treff på indikator

9.3 Funn indikator 3: «Mørk» AIS-aktivitet

Funn på indikator 3 er definert som at et fartøy ikke har rapport posisjon via AIS siste seks timer.

Rent teknisk var den opprinnelige idéen at det skulle etableres en tabell som kontinuerlig rapporterte siste posisjon til alle fiskefartøy. Basert på denne skulle det etableres en ny tabell som fortløpende skulle populeres med alle fartøy som ikke hadde rapportert posisjon siste seks timer. Denne analysemetoden er dessverre ikke mulig med det programvareoppsettet som benyttes, da programvaren ikke indekserer tabellen på en slik måte at denne type analyser kan gjennomføres.

Selv om det ikke er mulig automatisere denne analysemetoden, er det funksjonalitet i programvaren som gjør det mulig å få et delvis svar på denne indikatoren. Det er ikke mulig å gjennomføre dette som en såkalt «push query», altså en løpende spørring, men det kan gjennomføres som en såkalt «pull query». Det er en spørring som gjøres én gang, og deretter lukkes. Denne typen spørringer kan heller ikke gjøres mot en hel strøm eller tabell, da den må spesifiseres mot en verdi i tabellen, i dette tilfellet ble det gjort spørringer mot noen spesifikke MMSI.

Siden det ikke kan analyseres fortløpende over alle fartøy i datastrømmen må tilnærmingen endres og utvalget blir annerledes. Det er i teorien mulig å gjennomføre manuelle spørringer mot alle fiskefartøy i norske farvann, men dette vil vært svært tidkrevende og en ineffektiv måte å løse det på. Det som er interessant er å identifisere de fartøyene som har treff på flere enn én indikator. Derfor ble det naturlig å konsentrere seg om de fiskefartøyene som allerede var identifisert i arbeidet med indikator 1 og indikator 2. Dette utgjorde totalt åtte fartøy klassifisert som fiskefartøy.

MMSI	Fartøynavn	LAT	LON	SISTE observasjon	Minuter siden forrige obs	Treff på indikator
258216000	EIGENES	57.3086	8.0784	06-02-2021 20:54:22	552 (9t 12 min)	Ja
258654000	KINGS BAY	null	null	null	> 1080 min (18 timer)	Ja
273440730	KOTOYAVRI	null	null	null	> 1080 min (18 timer)	Ja

Tabell 8: Resultater fra indikator 3

Tabell 8 viser resultatene og viser at for tre av fartøyene har det gått lenger enn seks timer siden forrige posisjonsrapportering. Som tabellen viser resulterte ikke spørringene etter siste posisjonsoppdatering for fartøyene «Kings Bay» og «Kotoyavri» i noe svar i det hele tatt, derav null i flere kolonner. Basert på dette kan det konkluderes at fartøyene ikke har rapportert posisjon ila innhentingsfasen for denne indikatoren, dvs. at de ikke har rapportert posisjon via AIS siste 18 timer. Dette gjør at det ikke finnes koordinater lagret for siste posisjon til disse fartøyene, derfor vises de heller ikke kartet i figur 17.

Figur 17: Kartet viser siste rapporterte posisjon til fartøyene som har treff på indikator 3

9.4 Oppsummerte funn

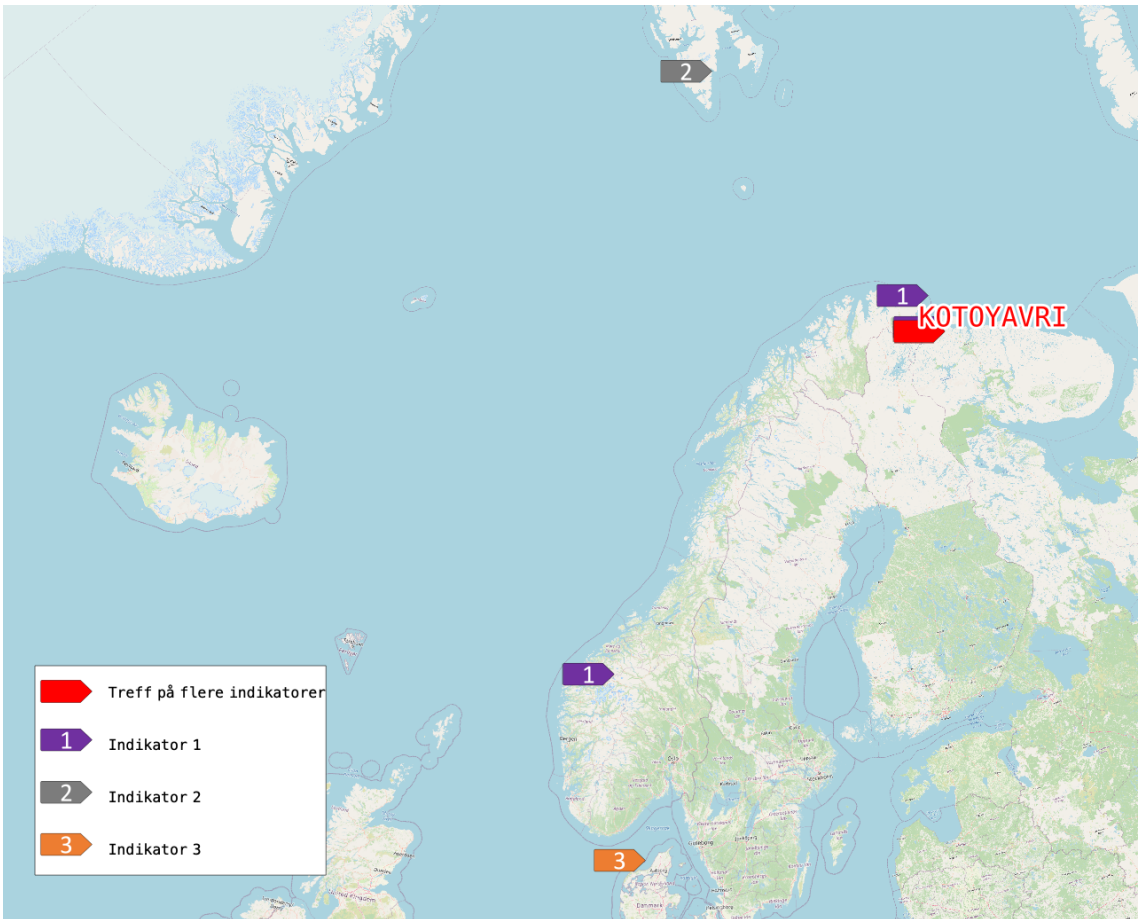
Siste steg i denne delen av analyseprosessen er å sette sammen resultatet fra alle delprosessen for å identifisere om det er noen av fiskefartøyene som gir utslag på flere av indikatorene. Som tidligere nevnt er ikke utslag på en indikator alene nok til å falsifisere hypotesene, men treff på flere indikatorer på samme fiskefartøy vil øke sannsynligheten for at fartøy driver UUU-fiske.

Fartøynavn	MMSI	Indikator 1	Indikator 2	Indikator 3	Score
KOTOYAVRI	273440730	1		1	**2**
EIGENES	258216000			1	1
MYS SLEPIKOVOSKOGO	273897000		1		1
ASPER Victoria	257390000	1			1
SENATOR	275171000	1			1
KINGS BAY	258654000			1	1

Tabell 9: Oppsummerte resultater fra analysearbeidet, som inkluderer alle fartøy som har utslag på en eller flere av indikatorer

I tabell 9 er alle fartøy som har treff på en eller fler indikatorer listet opp, disse funnene er også visualisert i figur 18. Det er verdt å merke seg at visualiseringen i figur 18 er noe upresis da tidspunktet ikke er det samme for alle fartøyene, men det er allikevel interessant å vise dette da det kan tjene som et eksempel på hvordan slik funn kunne vært visualisert.

Basert på funnene i denne studien er det kun ett fartøy som har treff på mer enn to indikatorer; det russiske fiskefartøyet «Kotoyavri». Fartøyet har både en historikk med UUU-fiske, samt at fartøyet operer uten AIS-transponderen påslått. Det skal understrekes at selv om fartøyet har utslag på to av indikatorene betyr ikke dette at fartøyet driver UUU-fiske. Allikevel betyr utslag på mer enn én indikator at det er *større sannsynlighet* for at fartøyet driver med UUU-fiske, gitt at analysene og indikatorene som er lagt til grunn stemmer.



Figur 18: Kart som viser hvor fartøyene som har treff på de ulike indikatorene er lokalisert

10 Diskusjon

I dette kapittelet vil eksperimentet evalueres. Det vil også drøftes hvorvidt artefaktbehovet, som beskrevet i kapittel 2, har blitt oppfylt. Evalueringen er strukturert rundt oppgavens forskningsspørsmål og tar for seg den metodiske og den tekniske tilnærmingen til stordatadrevet ACH. Videre diskuteres det i hvilken grad metoden gir merverdi og om metoden frigjør tid til å drive med mer dyptgående analyser. Deretter vurderes gyldigheten og påliteligheten til eksperimentet.

10.1 Forskningsspørsmål 1: Hvordan kan stordataanalyse og ACH integreres metodisk?

For å svare på forskningsspørsmålet vil dette delkapittelet struktureres i tre deler. Innledningsvis vil det diskuteres i hvilken grad stordatadrevet ACH er i samsvar med den opprinnelige metodikken til Heuer. Deretter vil det diskuteres hvorvidt problemstillingen med UUU-fiske i eksperimentet faktisk benytter ACH som metode på en hensiktsmessig måte. Avslutningsvis diskuteres det hvilke typer problemstillinger stordatadrevet ACH er egnet til å løse.

Heuer utviklet ACH som et støtteverktøy for å hjelpe analytikere med å vurdere ulike forklaringer på en problemstilling. Metoden legger opp til at alle tilgjengelige bevis skal vurderes opp mot de ulike hypotesene. Det er sentralt at man aktivt leter både etter bevis som både bekrefter og som avkrefter hypotesene (Heuer & Pherson, 2010, s. 161-162).

Stordatadrevet ACH har likheter med Heuers opprinnelige metode, men skiller seg klart når det kommer til hvilken type informasjon de behandler. I Heuers opprinnelige metode vil man bruke all tilgjengelig informasjon, inkludert antagelser, og også observasjoner man kan forvente å *ikke* se. Stordatadrevet ACH benytter seg av stordatakilder som er relativt strukturerte som datagrunnlag, og leter etter definerte indikatorer. Hvilke stordatakilder som er tilgjengelige vil være det som setter rammene for hva en kan analysere og strukturen på disse dataene vil definere hva du kan definere som indikatorer. Det kan argumenteres for at dette medfører at man får et mindre datatilfang, da langt fra all informasjon ligger tilgjengelig som informasjonsstrømmer eller databaser. På den annen side blir teknologien mer moden og infrastrukturen mer tilgjengelig nesten daglig, og det jobbes med å tilgjengeliggjøre data både i det private og det offentlige (Vivento & Kaupang, 2015). Dette gjør at datatilfanget og dermed analysemulighetene øker nesten daglig.

Sentralt i ACH er falsifisering av hypoteser, og det er et ideal å forsøke å falsifisere hypotesene fremfor å bekrefte dem (Popper, 2002). Samtidig er bekreftelse av enkeltindikatorer den mest effektive måten å besvare en indikator. Dette paradokset må forsøkes fanget opp ved formulering av

indikatorene. Med andre ord bør indikatorene som falsifiserer hypotesene faktisk kunne observeres (Hærens Våpenskole, 2021, s. 141; Heuer, 2005, s. 160). Dette paradokset blir også tydelig ved det utførte eksperimentet. Indikatorene fra C4ADS (Brush, 2019) som eksperimentet tar utgangspunkt i er utformet på en slik måte at de leter etter bekreftelse på hypotesen om at det foregår UUU-fiske. Metodisk er dette en utfordring, men samtidig er fraværet av svar på en indikator også en avkreftelse av den andre hypotesen, da hypotesene i eksperimentet er binære. Videre er en algoritme eller en datamaskin ikke er utsatt for kognitiv letthet (Kahneman, 2011) eller mentale snarveier. En datamaskin vil kun finne det den blir bedt om å finne, hverken mer eller mindre.

Det er derfor viktigere å identifisere gode indikatorer i stordatadrevet ACH enn i Heuers opprinnelige metode. Indikatorene må være definert slik at de kan identifiseres av en algoritme. Ettersom datamaskinen kun vil lete etter det du ber den om, er det helt sentralt at disse er presist definert. Selv små unøyaktigheter i beskrivelsen kan få store utslag. Søker algoritmene for bredt får man for mange svar, og hvis de defineres for smalt får man ingen svar. Det vil derfor være helt sentralt å benytte seg av fagekspertise under utarbeidelsen av indikatorer for å få mest mulig nøyaktig og presis nedbryting av det fenomenet man ønsker å observere. Jo mer presist man kan beskrive fenomenet man skal finne, jo mer presis vil analysen bli.

Et annet aspekt som det må tas høyde for er at det er utfordrende å vurdere en indikators påvirkning og at vi er særlig utsatt for bias når vi gjør det (Dawes, 2001). Denne utfordringen kan gjøre seg gjeldende hvis man ønsker å vekte de ulike indikatorens påvirkning på den totale inkonsistensscoren. Metoden åpner for å kunne vekte enkelte indikatorer mer eller mindre enn andre, men å gjøre dette på en god måte krever fagekspertise tilgjengelig som kan være med på å vurdere den enkelte indikators påvirkning. Hvis dette ikke tas høyde for når det utvikles indikatorer vil man kunne få en meget stor systematisk skjevhet i analysene.

I eksperimentet opereres det med to hypoteser, enten driver et fartøy med UUU-fiske eller så gjør det ikke det. Disse hypotesene er gjensidig utelukkende slik metoden oppfordrer til (Heuer & Pherson, 2010, s. 161-162); samtidig er hypotesene binære, det er ja eller nei. Forsvarets etterretningsdoktrine (Forsvarssjefen, 2021, s. 69) anbefaler hypotesebasert tilnærming først når problemstillinger har fler enn to utfall eller forklaringer. Dette er et argument for at metoden er tiltenkt å brukes for å vekte mellom fler hypoteser enn kun to, og at metoden egner seg enda bedre når det er fler hypoteser. Det kan videre hevdes at det finnes enklere måter å håndtere hypoteser på som har svaret ja eller nei. Derav kan man spørre seg om det er hensiktsmessig å benytte seg av ACH i det hele tatt. Selv om det i eksperimentet er en binær hypotese muliggjør teknologien å teste

hypoteser ned på fartøynivå. Dette betyr at man samtidig tester hypotesene og indikatorene på over 500 fiskefartøy[5]. Denne hypotesetestingen foregår kontinuerlig og gir et betydelige nivå av kompleksitet selv om hypotesen som testes er binær. Dette tilfører en ny dimensjon til metoden.

Samtidig argumenter Heuer for at metoden også har sin nytte når det kommer til konkrete problemstillinger. Det trekkes frem at metoden egner seg godt til tekniske problemstillinger, slik som *«Hvilket våpensystem importeres denne spesifikke delen til?»* (Heuer & Pherson, 2010, s. 161). I et slik scenario vil man sammenligne ulike bevis med kunnskap man har om våpensystemet. Dette er ikke helt ulikt problemstillingen som det arbeides med i dette eksperimentet.

Flere sentrale etterretningsteoretikere anvender begrepene «*secrets, mysteries and complexities»* (Hatlebrekke, 2019; Omand, 2010). *«Secrets»* er data eller informasjon som eksisterer, men som en aktør ønsker å holde hemmelig, for eksempel hvor militære våpenplattformer er og hva de gjør. Den andre typen er *«mysteries»* og omhandler informasjon om utfall av fremtidige hendelser, intensjoner som ennå ikke har kommet til uttrykk, eller beslutninger som ennå ikke er fattet. Den tredje og siste typen er *«complexities»,* som tar høyde for at egne handlinger kan påvirke andre aktørers handlinger (Forsvarssjefen, 2021, s. 27-29).

Metoden slik den er definert nå vil være best egnet til å løse problemer som faller inn under kategorien *«secrets».* Dette er ofte problemstillinger som kan brytes ned i tid og rom og hvor mye av analysene vil omhandle å identifisere mønstre eller avvik fra normalen. Det gjennomførte eksperimentet har vist at stordatadrevet ACH vil fungere godt til nettopp denne type bruk. Dette underbygges også av Borg (2017) som poengter at datadrevne analyser ikke er egnet for alle typer problemstillinger, men kan ha sin nytte når man leter etter såkalte *«secrets».* Dette understreker at stordatadrevet ACH ikke nødvendigvis er en løsning på alle problemstillinger. For mer komplekse problemstillinger hvor det handler om å avdekke intensjoner, eller hvor aktørene opererer i større grad opererer i det skjulte som f.eks. ved COIN[6]-operasjoner, eller i de tilfeller hvor det er utfordrende å uttrykke konkrete indikatorer, så har metoden visse svakheter.

[5] I følge Statistisk sentralbyrå (2019) var det i 2019 registrert 470 norske fiskefartøy over 15 meter. Hvis det også inkluderes utenlandske fartøy over 15 meter som opererer i norske farvann er totalen mer enn 500 fiskefartøy.
[6] COIN er en forkortelse for *«counterinsurgency»*. På norsk opprørsbekjempelse.

10.2 Forskningsspørsmål 2: Hvordan kan stordataanalyse og ACH integreres teknisk?

Den foreslåtte tekniske løsningen presenteres i kapittel 6. Her redegjøres det for en infrastruktur som består av fire ulike komponenter. Disse komponentene sørget for automatisering av databehandlingen i noen av stegene i stordatadrevet ACH ifm. gjennomføringen av eksperimentet presentert i kapittel 8.

For å diskutere forskningsspørsmål 2 er det verdt å hente frem artefaktbehovet definert i kapittel 2. Dette ble definert som: «*Et verktøy som i sanntid analyserer en kontinuerlig stordatastrøm, og fortløpende sjekker om disse dataene tilfredsstiller et sett med forhåndsdefinerte indikatorer og hypoteser, og svarer dette til bruker fortløpende.»* Den videre diskusjonen vil struktureres rundt to hovedpunkter, for det første vil det diskuteres hvorvidt verktøyet evnet å analysere en kontinuerlig stordatastrøm, og for det andre vil det diskuteres hvordan dette ble presentert til brukeren.

Stordatainfrastrukturen har overordnet fungert godt og komponentene har i samspill sørget for at indikatorene har blitt besvart. Som presentert i kapittel 9 viser resultatet av analysen at det har blitt identifisert *ett* fiskefartøy som det er økt sannsynlighet for at driver med UUU-fiske. Stordatainfrastrukturen evnet å gi disse svarene i nær sanntid og håndterte sånn sett datamengden på en god måte. Selv om sluttresultatet viser at metoden fungerer, er ikke artefaktbehovet fullstendig oppfylt da det mangler noe funksjonalitet for at bruker skal kunne få sine svar fortløpende. Noen av disse utfordringene er tilknyttet analyseverktøyet, ksqlDB.

ksqlDB er godt egnet til å gjøre enklere analyseoppgaver som å splitte datastrømmer basert på attributter eller utføre statistikkberegninger og enkle analyseoppgaver. Slike operasjoner er enkle å sette opp, analysene går hurtig og resultatene publiseres i nær sanntid. Begrensningene i programvaren blir tydeligst når det skal analyseres store mengder romlige data, altså data med x- og y-koordinater. Eksempelvis krevde regneoperasjonen for å identifisere om fartøy var nærmere hverandre enn 500 meter, nesten all tilgjengelig lagringsplass og prosessorkraft. Som et resultat ble alle andre programmer og prosesser som kjørte på samme maskin svært lite responsive. Det gjorde det svært tidkrevende å sjekke om man hadde fått resultater. Det medførte også maskinhavari etter 36 timer. En måte å mitigere denne utfordringen hadde vært å distribuere denne oppgaven utover en maskinklynge på flere maskiner, men for å ikke komplisere eksperimentet ble ksqlDB kjørt på en maskin.

ksqlDB ble valgt fordi det er et av få stordataanalyseverktøy som ikke krever at brukeren kan programmering. Dette hensynet har veid tungt og har vært nødvendig for i det hele tatt å få gjennomført eksperimentet. Dette gjorde at andre muligheter, som bruken av resonneringsrammverket LARS (Beck, Dao-Tran, Either & Fink, 2015) eller geografiske analyser ved hjelp av rammeverket Apache Spark (Corizzo, Ceci & Japkowicz, 2019), ble valgt vekk da brukerterskelen ble vurdert til å være for høy sett opp mot tid tilgjengelig for oppgaven.

Brukerterskelen for programvaren i stordatainfrastrukturen vurderes å være relativt høy for den normale databruker. Eksempelvis er det en forutsetning å være komfortabel med å jobbe i kommandolinje da flere av programmene ikke har egne grafiske grensesnitt. Videre er oppsett av en stordatainfrastruktur tidkrevende, og det gikk mye tid innledningsvis i eksperimentperioden til å få de ulike komponentene til å snakke med hverandre og få løsningen til å fungere som ønsket. Utfordringer som eksempelvis å holde kontroll på hvilke programvareversjoner som er kompatible med hverandre og korrekt konfigurasjon av Docker er ingen triviell oppgave. Opprinnelige var det planlagt å benytte enda to komponenter (GeoMesa og GeoServer) for å visualisere svarene i sanntid, men grunnet tekniske utfordringer med disse ble det bestemt å droppe dem og heller gjennomføre et eksperiment med enklere oppsett. Det ble derfor valgt å lese ut dataene manuelt fra Kafka og legge dem manuelt over i QGIS for visualisering.

Et tiltak som kunne møtt flere av de tekniske utfordringene knyttet til oppsett, drift og valg av programvare ifm. gjennomføringen av eksperimentet hadde vært et samarbeid med en dataingeniør. De fleste av de tekniske utfordringene kunne antageligvis vært løst på en brøkdel av tiden og med et bedre resultat ved å jobbe tett på en dataingeniør med nødvendig fagkunnskap til å etablere og drifte stordatainfrastrukturen. Dette hadde også medført at man hadde hatt mulighet til å velge programvare som eksempelvis hadde analysert romlige data på en mer effektiv måte. Det virker derfor rimelig å anta at et suksesskriterium for å lykkes med å utvikle gode stordataanalyseløsninger er tett samarbeid mellom brukere, i dette tilfellet etterretningsanalytikere, utviklere og dataingeniører. Det at man forstår hverandres utfordringer vil gjøre utvikling og implementering smidigere, og samtidig øke sannsynligheten for at man ender med ett verktøy som bruker vil ha, og som dataingeniørene kan sette opp og drifte.

Oppsummert har den tekniske løsningen delvis oppfylt artefaktbehovet. Stordatainfrastrukturen tilfredsstiller første del av behovet ved at det er utviklet «*et verktøy som i sanntid analyserer en kontinuerlig stordatastrøm, og fortløpende sjekker om disse dataene tilfredsstiller et sett med forhåndsdefinerte indikatorer og hypoteser ...».* Selv om prosessen ikke er helautomatisert er det så

nært at vi kan si at denne delen av artefaktbehovet er oppfylt. Hva gjelder den andre delen av artefaktbehovet «… *og svarer dette til bruker fortløpende»* har det blitt vist en løsning som svarer til bruker, men ikke fortløpende. Utfordringene knyttet til denne delen av artefaktbehovet er først og fremst knyttet til begrenset kunnskap og erfaring med stordatainfrastruktur. Ved enten å ha mer tid tilgjengelig for å gjennomføre eksperimentet eller ved å samarbeide med dataingeniør(er) hadde dette trolig vært løsbart.

10.3 Forskningsspørsmål 3: Tilfører metoden merverdi og vil den den gjøre at analytikeren kan bruke mer tid på analyse?

Utgangspunktet for problemstillingen i oppgaven var å studere om kombinasjonen av stordata og etterretningsanalyse kunne gi mer effektiv etterretningsanalyse. Det må altså gi en merverdi å sette opp og etablere en slik løsning. Merverdi i denne sammenheng defineres som at stordatadrevet ACH har medført at svarene på analysene har en større verdi enn de dataene som ble hentet inn i systemet, sammenholdt med tiden vi har lagt ned i arbeidet.

Jeg vil hevde at metoden tilfører merverdi til analyseprosessen all den tid den faktisk gir svar som fremstår som troverdige. Stordatadrevet ACH gir mulighet til å analysere stordata med en presisjon og en hurtighet som ikke er mulig med manuelle analysemetoder. Metoden gjør det mulig å søke gjennom enorme datamengder og detektere avvik som ellers er vanskelig eller umulig å oppdage. Mens mennesker har utfordringer med å koble bevis til mer enn tre hypoteser uten hjelpemidler (Heuer & Pherson, 2010) har eksperimentet vist at verktøyet evner å analysere mer enn 5 millioner meldinger og kontinuerlig sjekke om dette gir treff på noen av indikatorene på mer enn 500 unike fartøy samtidig. Det å holde styr på alle 500 fartøyene i sanntid og kontinuerlig monitorere for alle mulige indikatorer er utenkelig å gjøre manuelt. Stordatadrevet ACH gjør det derfor mulig å overvåke store områder med relativt liten bemanning, gitt at man investerer tid i det initielle arbeidet med utvikling av gode indikatorer og algoritmer. Verktøyet har et potensial til å gjøre det langt lettere for analytikeren å håndtere et voksende hav av informasjon, trekke ut det som er relevant, og få denne informasjonen presentert i strukturert form.

Samtidig skal det ikke legges skjul på at det dette kommer med en kostnad, både tidsmessig og økonomisk, og at det er et spørsmål hvor mye tid man vil spare. Spesielt vil det være en stor kostnad innledningsvis fordi det kreves et betydelig arbeid for å etablere en slik løsning da det er viktig med grundige analyser i forkant for å definere behov og kravspesifikasjoner. Videre er det viktig med tett samarbeid underveis i utvikling og implementering av en slik løsning for å kvalitetssikre at arbeidet går i rett retning, samt kvalitetssikre at løsningen løser de oppdrag den skal når den først har

kommet i drift. Selv om programvaren er gratis krever det altså et betydelig antall arbeidstimer både fra de som har behov for verktøyet og de som skal utvikle verktøyet.

Grunnet denne tidsbruken for implementering og etablering vil slike løsninger ikke være egnet til alle problemstillinger. Generelt vil det å benytte seg av metoder som stordatadrevet ACH egne seg til problemstillinger som skal overvåkes over noe tid slik at tiden som brukes til etablering står i forhold til hva analysene gir. Det vil eksempelvis være hensiktsmessig å etablere en løsning som skal overvåke UUU-fiske i norske farvann. Problemstillingen knyttet til UUU-fiske vil være relevant for norske myndigheter i overskuelige fremtid og i tillegg vil datatilgangen være god i lang tid fremover (Kystverket, 2021). Andre mer kortsiktige problemstillinger vil ikke være egnet.

I tillegg er det verdt å understreke at stordatadrevet ACH ikke er tiltenkt å være et selvstendig analysesystem. Til det er analysene for snevre og for spesifikke, og det er en utfordring at den kun evner å tolke strukturert informasjon. Et slikt verktøy vil kunne fungere som et støtteverktøy som gjør at visse deler av analyseprosessen kan gjennomføres raskere og med høyere presisjon. Løsninger basert på stordataanalyse kan ikke i dag svare på alle spørsmål, men utviklingen innen fagfeltet går med stor hastighet og det vanskelig å spå hvordan framtiden blir. Spørsmålet blir derfor om morgendagens analytikere vil være i stand til å konfrontere fremtiden uten å benytte seg av slik teknologi.

10.4 Gyldighet og pålitelighet

I praksis er absolutt gyldighet og pålitelighet uoppnåelige idealer, men det er viktig at man etterstreber dette i så stor grad som mulig (Stølen, 2019, s. 122). Gyldighet (også kjent som validitet) forteller oss om man kan trekke relevante og meningsfulle slutninger fra undersøkelsen (Creswell, 2014, s. 250). Gyldigheten forteller oss i hvor stor grad benyttede data er gyldige for problemstillingen studien tar for seg (Busch, 2013, s. 62). Eksperimentet i denne oppgaven er altså gyldig hvis det evaluerer det det er ment å evaluere (Stølen, 2019, s. 121). Med andre ord hvis eksperimentet faktisk tester om stordatadrevet ACH er en hensiktsmessig måte å kombinere stordataanalyse og etterretningsanalyse.

Pålitelighet (også kjent som reliabilitet) er knyttet til målekvalitet og om vi kan stole på dataene som er kartlagt (Busch, 2013, s. 62). Påliteligheten kan altså si oss noe om innsamling av data er gjort på en pålitelig måte, og om en re-test av datagrunnlaget hadde gitt høy grad av korrelasjon (Creswell, 2014, s. 247). Man kan si at eksperimentet er pålitelig hvis det kan gjentas og gi tilnærmet samme resultat hver gang det gjentas (Stølen, 2019, s. 121).

Stordatainfrastrukturen ble evaluert ved hjelp av et felteksperiment noe som betyr at evalueringen skal foregå i naturlige omgivelser. Ved å velge å gjennomføre et felteksperiment tas et valg om å ofre presisjon for realisme, altså har det blitt ofret pålitelighet for gyldighet. Det ble derfor gjort et valg om å benytte en sanntidsdatastrøm i et tilfeldig tidsvindu som datagrunnlag for eksperimentet. Dette betyr at det ikke er mulig å gjenskape akkurat det samme eksperimentet med de samme dataene, men et lignende eksperiment kan gjennomføres når som helst. Resultatene vil ikke være identiske, men man kan forvente å få lignende resultater. Dette underbygges av funn jeg gjorde under uttesting av systemet, i forkant av selve innhentingsfasen. Analyseresultatene jeg fikk da var sammenlignbare med de jeg samlet inn under innhentingsfasen, noe som antyder at påliteligheten har vært akseptabel.

Videre er det generelt endel støy i AIS-data. Eksempelvis forekommer det duplisering av enkeltmeldinger, meldinger med feil verdier og andre feilkilder som drøftet i kapittel 7. Dette vil kunne føre til unøyaktigheter i analysearbeidet. Spesielt sårbar er algoritmene knyttet til analysen for å identifisere omlasting til sjøs, og selv små feil her kan potensielt skape unøyaktige sluttresultater. Et tilrettelagt datasett som hadde vært «ryddet i» og klargjort kunne gitt høyere presisjon på analysene, men da hadde det ikke vært et felteksperiment og realismen i eksperimentet ville blitt redusert. I tillegg er det tatt høyde for disse feilkildene i utformingen av algoritmen for å redusere effekten slik støy har på analyseresultatene.

Selve innhentingsfasen kunne vart lenger og med det hatt et mer omfattende datagrunnlag, for å få flere treff på de ulike indikatorene, samt at man i større grad hadde fått testet hvor robust stordatainfrastrukturen er. Dette lot seg ikke gjøre da innhenting og analyse på datagrunnlaget gjorde at harddisken ble full på 36 timer, og datainfrastrukturen stanset som en følge av dette. På tross av at testdatasettet har noe kort varighet er allikevel resultatene fra analysene gode nok til å gi svar på alle indikatorene. Hva gjelder påliteligheten for hele eksperimentet vurderes den som god, da man kunne forventet å få lignende resultat hvis man hadde gjennomført samme analyse i dag.

For å øke både pålitelighet og gyldighet kunne den foreslåtte metoden vært testet av en uavhengig testgruppe. Dette kunne bedre kvalitetssikret om metoden er gjennomførbar samt om den er godt nok beskrevet. Jeg med min erfaring har ett utgangspunkt, og det som er selvsagt og innlysende for meg er det ikke sikkert at er det for andre. På den annen side har det blitt fokusert på å dokumentere prosessen og gjennom dette sørge for at metodikken skal være etterprøvbar.

Det å ha gode indikatorer og at disse indikatorene måler det de faktisk skal er sentralt i eksperimentet og i analysemetoden. Indikatorene til eksperimentet har basert seg på studien til Brush (2019). Studien fremstår troverdig, er godt referert og viser til eksempler på hvor de har lykkes med sin analysemetodikk, men jeg har ikke har funnet noen studier som nyanserer eller problematiserer denne. Studien har et globalt fokus og med tanke på at eksperimentet kun ser på norske farvann ville det styrket indikatorene, og eksperimentet, med studier som hadde sett på regionale forskjeller, og fokus på unike utfordringer i norske farvann. Selv om det ikke finnes studier som ettergår Brush sine funn direkte, bygges troverdigheten opp av at lignende indikatorer benyttes i andre rapporter. Eksempelvis setter Fenton (2020) fokus på fartøy som skrur av AIS med hensikt, mens Kroodsma (2017) setter fokus på omlastingsaktivitet i sin rapport.

11 Konklusjon

Den teknologiske utviklingen innenfor databehandling og stordatasystemer skjer i et stort tempo og det hevdes at denne utviklingen vil ha stor og økende betydning for militære operasjoner framover. Denne utviklingen vil gjøre det mulig å overføre, analysere og sammenstille informasjon raskere og mer effektivt, og gjennom dette kunne tolke enorme datamengder for å etablere et overlegent situasjonsbilde. Et situasjonsbilde består av mange ulike deler, men en sentral del av et situasjonsbilde bygges ved hjelp av etterretningsanalyse.

Utgangspunktet for denne oppgaven var en observasjon av at stordata- og etterretningsanalyse har en del fellestrekk, og at stordataanalyse derfor kan ha et potensiale for å bidra til mer effektiv etterretningsanalyse. Målsetningen ble derfor å utforske om, og på hvilken måte stordataanalyse kan understøtte etterretningsanalyse. For å gjøre dette har det blitt foreslått en metode som kombinerer kjente analysemetoder for stordata og etterretningsanalyse. Metoden har fått navnet *stordatadrevet ACH*. Metodens ambisjon er å muliggjøre søk gjennom store datavolum med høy presisjon og med en tydelig metodisk forankring i etterretningsanalyse.

Stordatadrevet ACH benytter stordataanalyse til å avdekke mønstre og foreslå konklusjoner, mens ACH benyttes som et rammeverk for å utvikle hypoteser, vurdere kontekst og ta de endelige beslutningene. Ved å kombinere to forskjellige metoder er hensikten å prøve å utnytte menneskets kreativitet og datamaskinenes presisjon og nøyaktighet. Metoden legger også opp til at dette skal svares i sanntid til brukeren.

For å kunne gjennomføre stordataanalyser er man avhengig av en stordatainfrastruktur. Det finnes i dag ingen programvare som løser alle problemer noe som gjør at design av stordatainfrastruktur i stor grad kan betraktes som skreddersøm. Valget av programvare vil basere seg på en analyse av hvilke datakilder som skal benyttes og tiltenkte analyseoppgaver.

For å kunne måle om stordatadrevet ACH fungerer ble det definert et såkalt artefaktbehov. Dette er uttrykt på følgende måte: *«Et verktøy som i sanntid analyserer en kontinuerlig stordatastrøm, og fortløpende sjekker om disse dataene tilfredsstiller et sett med forhåndsdefinerte indikatorer og hypoteser, og svarer dette til bruker fortløpende.» Verktøy* referer her til kombinasjonen av metode og stordatainfrastruktur. For å evaluere om artefaktbehovet ble nådd ble det gjennomført et eksperiment.

I eksperimentet ble verktøyet testet ved å søke etter svar på en hypotese om hvorvidt det foregår ulovlig, urapportert eller uregulert (UUU) fiske i norske farvann. Stordatadrevet ACH ble benyttet som analytisk rammeverk for å analysere problemstillingen, velge stordatainfrastruktur, lage algoritmer for å svare på problemstillingen og til slutt hente inn svar på problemstillingen. Resultatet av prosessen var at det ble identifisert *ett* fiskefartøy vi kan hevde at det er økt sannsynlighet for at driver med UUU-fiske.

Verktøyet som ble utviklet har delvis tilfredsstilt artefaktbehovet. Løsningen evner å analysere datastrømmen og besvare hypotesene, men slik stordatainfrastrukturen og verktøyet er satt opp nå manglet det en god måte å presentere dataene på i sanntid. Dette kunne vært løst ved å inkludere flere eller andre komponenter i stordatainfrastrukturen, men grunnet tid tilgjengelig ble det prioritert å få gjennomført eksperimentet fremfor en optimal stordatainfrastruktur.

Et sentralt funn er at det er tidkrevende og komplisert å etablere stordatainfrastruktur. Problemanalyser, valg av komponenter, oppsett og konfigurasjon tar tid. I tillegg er det komplekse systemer som krever spesifikk kompetanse både for å sette opp, konfigurere og drifte. Videre er ikke stordataløsninger egnet til å løse alle typer problemstillinger. For det første må man ha en stordatakilde, for det andre må problemstillingen være av en slik art at den kan brytes ned til spesifikke indikatorer som systemet kan lete etter. I sum medfører dette at stordataløsninger passer best til problemstillinger som har et langsiktig perspektiv. Med andre ord, for at løsningen skal gi merverdi må innsatsen med å etablere løsningen stå i forhold til gevinsten man får. Samtidig vurderes tett samarbeid mellom etterretningsanalytikere, utviklere og dataingeniører å være et suksesskriterium for å lykkes med å utvikle gode stordataanalyseløsninger. Dette vil øke sannsynligheten for at man ender med et verktøy som analytikerne vil ha, og som dataingeniørene kan sette opp og drifte.

Bruken av stordatadrevet ACH vil for visse problemstillinger frigjøre kapasitet hos analytikeren. Dette er tid som kan fokuseres på videre analyser og kontekstualisering av informasjon, og gjennom dette øke situasjonsforståelse og evne til prediksjon, altså i kort gi bedre etterretninger. Dette gjør at merverdien av metoden alltid må sees i sammenheng med andre analyseverktøy, og ikke alene stå som beslutningsgrunnlag.

Studien viser at stordatadrevet ACH er et kraftig verktøy når det anvendes riktig. Den endelige konklusjonen er derfor at stordatadrevet ACH kan øke omfanget og anvendeligheten til ACH spesielt og til etterretningsanalyse generelt, men dette forutsetter at det benyttes på rett problemstillinger.

Det må ses på som et supplement til eksisterende prosesser og systemer, ikke en erstatning. Det anbefales derfor å forske videre på problemstillinger knyttet til denne tematikken, for å vurdere hvorvidt stordataanalyse også i andre sammenhenger kan bidra til økt operativitet.

11.1 Anbefalt videre forskning

Oppgaven tar for seg en problemstilling som i liten grad er studert, og det står igjen ubesvarte spørsmål som bør studeres videre. Det anbefales at fremtidige studier ser på militær bruk av stordatadrevet ACH. Eksempelvis vil det være interessant å forske på bruken av stordatadrevet ACH til å monitorere en aktørs handlemåter. På operasjonelt og taktisk nivå vil etterretningsstaben utarbeide minst to fiendtlige handlemåter, gjerne flere. Disse handlemåtene kan betraktes som hypoteser. I tilknytning til handlemåtene skal det utarbeides en indikatorliste. Det utarbeides også et såkalt mulighetsoverlegg som sier noen om hvilke av motstandernes enheter vi forventer å se hvor, og når vi forventer å se dem der. Alle disse nevnte delproduktene kan brukes som utgangspunkt for å teste hvordan stordatadrevet ACH ville fungert i en mer operativ setting.

Det anbefales også at fremtidige studier ser på hvilke mulige stordatakilder som finnes tilgjengelig innad i Forsvaret og hvordan disse kan tilgjengeliggjøres for stordataanalyser. Videre hadde det vært interessant å se på hvordan disse stordatakildene og analyseverktøy kunne blitt integrert på operative kommando og kontroll (K2)-plattformer i Forsvaret for å understøtte operative behov.

Til slutt hadde det også vært interessant å forske videre se på hvordan den spesifikke problemstillingen knyttet til UUU-fiske kunne vært utviklet videre, herunder modellere de resterende identifiserte indikatorene, se på hvilke andre datakilder som kunne vært benyttet for å utvide analysemulighetene, og integrasjon av slikt verktøy hos Kystverket eller ved Forsvarets Operative Hovedkvarter.

Litteraturliste

Artner, S. J., Girven, R. & Bruce, J. B. (2016). *Assessing the value of structured analytic techniques in the US intelligence community* Rand Corporation.

Balazka, D. & Rodighiero, D. (2020). Big Data and the Little Big Bang: An Epistemological (R)evolution. *Frontiers in Big Data, 3*(31). https://doi.org/10.3389/fdata.2020.00031

Beadle, A. W., Diesen, S., Nyhamar, T. & Bostad, E. K. (2019). *Globale trender mot 2040 : : et oppdatert fremtidsbilde*. Kjeller: Forsvarets forskningsinstitutt FFI.

Beck, H., Dao-Tran, M., Either, T. & Fink, M. (2015). LARS: A Logic-based Framework for Analyzing Reasoning over Streams.

Bergander, K. & Johnsen, B. (2006). Vitenskap og metode.

Bokmålsordboka. (2021). Teknologi. Hentet 3. mars 2021 fra https://ordbok.uib.no/perl/ordbok.cgi?OPP=teknologi&ant_bokmaal=5&ant_nynorsk=5&begge=+&ordbok=begge

Borg, L. C. (2017). Improving intelligence analysis: Harnessing intuition and reducing biases by means of structured methodology. *The International Journal of Intelligence, Security, and Public Affairs, 19*(1), 2-22.

Brantly, A. F. (2018). When everything becomes intelligence: machine learning and the connected world. *INTELLIGENCE AND NATIONAL SECURITY, 33*(4), 562-573. https://doi.org/https://doi.org/10.1080/02684527.2018.1452555

Brasfield, A. D. (2009). *Forecasting Accuracy and Cognitive Bias in the Analysis of Competing Hypotheses* Mercyhurst College.

Brush, A. (2019). *Strings Attached: Exploring onshore networks behind illegal, unreported, and unregulated fishing*. Center for Advanced Defense Studies (C4ADS). Hentet fra https://static1.squarespace.com/static/566ef8b4d8af107232d5358a/t/5d7022301845f300016ee532/1567629912450/Strings+Attached.pdf

Busch, T. (2013). *Akademisk skriving for bachelor-og masterstudenter* Fagbokforl.

Central Intelligence Agency. (2009). A tradecraft primer: Structured analytic techniques for improving intelligence analysis. *CIA Center for the study of intelligence*. Hentet fra https://www.cia.gov/static/955180a45afe3f5013772c313b16face/Tradecraft-Primer-apr09.pdf

Corizzo, R., Ceci, M. & Japkowicz, N. (2019). Anomaly detection and repair for accurate predictions in geo-distributed big data. *Big Data Research, 16*, 18-35.

Coulthart, S. (2016). Why do analysts use structured analytic techniques? An in-depth study of an American intelligence agency. *INTELLIGENCE AND NATIONAL SECURITY, 31*(7), 933-948. https://doi.org/10.1080/02684527.2016.1140327

Creswell, J. W. (2014). *Research design : qualitative, quantitative, and mixed methods approaches* (4th ed.; International student ed. utg.). Los Angeles, Calif: SAGE.

Dawes, R. (2001). *Everyday Irrationality: How Pseudo-scientists, Lunatics, and the Rest of Us Systematically Fail to Think Rationally* Westview Press.

Department Of Defense. (2013). Joint Publication 2-0 Joint Intelligence. Hentet fra https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp2_0.pdf

Dhami, M. K., Belton, I. K. & Mandel, D. R. J. A. C. P. (2019). The “analysis of competing hypotheses” in intelligence analysis*, 33*(6), 1080-1090.

Docker. (2021). Docker overview. Hentet 21. mars 2021 fra https://docs.docker.com/get-started/overview/

Dvergsdal, H. & Elster, A. C. (2019). Stordata. I *Store Norske Leksikon - SNL.no*. Hentet 31. mars 2021 fra https://snl.no/stordata

Edgar, T. H. (2017). *Beyond snowden: privacy, mass surveillance, and the struggle to reform the NSA* Brookings Institution Press.

Emani, C. K., Cullot, N. & Nicolle, C. (2015). Understandable big data: a survey. *Computer science review, 17*, 70-81.
Eriksen, T., Greidanus, H. & Delaney, C. (2018). Metrics and provider-based results for completeness and temporal resolution of satellite-based AIS services. *Marine Policy, 93*, 80-92.
Etterretningstjenesten. (2013). *Etterretningsdoktrinen*.
Fenton, A. (2020, 27/01/2020). Shining a Light on Ships that Go Dark. Hentet fra https://wnwd.com/blog/shining-a-light-on-ships-that-go-dark/
Fiskeridirektoratet. (2020). Norwegian black list. Hentet 19. januar 2021 fra https://www.fiskeridir.no/Yrkesfiske/Kontroll/Ulovleg-fiske/Norwegian-black-list
Folker, R. D. (2000). *Intelligence analysis in theater joint intelligence centers: An experiment in applying structured methods* Joint Military Intelligence College.
Forsvaret. (2019). *Forsvarets fellesoperative doktrine* Forsvarsstaben.
Forsvarets forskningsinstitutt. (2019). *Forsvarsteknologiske trender : en overordnet analyse av teknologiens betydning for et effektivt og relevant forsvar*. Kjeller: Forsvarets forskningsinstitutt FFI.
Forsvarsdepartementet. (2019). *Prop 62 S - Vilje til beredskap – evne til forsvar Langtidsplan for forsvarssektoren 21-24*. Hentet fra https://www.regjeringen.no/contentassets/b43ae5a187034670adc96a83fbf79651/no/pdfs/prp201920200062000dddpdfs.pdf
Forsvarsdepartementet. (2020, 10. november 2020). Utsetter ikrafttredelse av to kapitler i e-loven. Hentet fra https://www.regjeringen.no/no/aktuelt/elovkap7og8/id2784463/
Forsvarssjefen. (2021). *Forsvarets etterretningsdoktrine*.
Froelich, P. (2020, 22. februar 2020). Elon Musk calls for regulations on artificial intelligence. *New York Post*. Hentet fra https://nypost.com/2020/02/22/elon-musk-calls-for-regulations-on-artificial-intelligence/
Gartner. (2012). Big Data. I *Gartner IT Glossary*. Hentet 31. mars 2021 fra https://www.gartner.com/en/information-technology/glossary/big-data
Gill, P. & Phythian, M. (2016). What Is Intelligence Studies? *The International Journal of Intelligence, Security, and Public Affairs, 18*(1), 5-19. https://doi.org/10.1080/23800992.2016.1150679
Grønmo, S. (2020). bias i forskning. I *Store norske leksikon*. Hentet fra https://snl.no/bias_i_forskning
Hærens Våpenskole. (2021). *Stabshåndbok for Hæren: Plan- og beslutningsprosessen* Hærens våpenskole.
Halvorsen, J. & Hansen, B. J. (2020). *Exploring data reuse using a big data infrastructure* (8246433040). Hentet fra https://www.ffi.no/en/publications-archive/exploring-data-reuse-using-a-big-data-infrastructure
Hatlebrekke, K. A. (2019). *The Problem of Secret Intelligence* Edinburgh University Press.
Haugom, L., Hemmingby, C. & Pedersen, T. (2019). Å analysere med kunstig intelligens. I S. Stenslie, L. Haugom & B. H. Vaage (Red.), *Etterretningsanalyse i den digitale tid - en innføring* (1. utgave. utg., s. 85-106). Bergen: Fagbokforlaget.
Heuer, R. J. (1999). *Psychology of intelligence analysis* Center for the Study of Intelligence.
Heuer, R. J. (2005). How does analysis of competing hypotheses (ACH) improve intelligence analysis.
Heuer, R. J. & Pherson, R. H. (2010). *Structured analytic techniques for intelligence analysis* CQ Press.
Hofstad, K. (2018). Breddegrad. I *Store Norske Leksikon*. Hentet 7. mars 2021 fra https://snl.no/breddegrad
IMO. (2015). A29/Res.1106. Revised guidelines for the onboard operational use of shipborne automatic identification systems (AIS).
Jafarpour, H., Desai, R. & Guy, D. (2019). KSQL: Streaming SQL Engine for Apache Kafka. *EDBT* (s. 524-533).
Johnsrud, H. J., Pedersen, K. J. & Gravir, G. (2014). *Vedlegg til SJØSIKKERHETSANALYSEN 2014 - Analyse av sannsynligheten for ulykker med tap av menneskeliv og akutt forurensning fra*

skipstrafikk i norske farvann. Hentet fra https://www.kystverket.no/contentassets/f056df3c875140aa98ef49a25cc082c6/4b_vedlegg-til-sannsynlighetsanalyse-for-2013-dagens-risiko.pdf

Jones, N. (2018). Critical epistemology for analysis of competing hypotheses. *INTELLIGENCE AND NATIONAL SECURITY, 33*(2), 273-289.

Kahneman, D. (2011). *Thinking, fast and slow* Macmillan.

Kent, S. (1949). *Strategic intelligence for American world policy* Princeton University Press.

Kjerstad, N. (2020). AIS. I *Store Norske Leksikon - SNL.NO*. Hentet 11. januar 2021 fra https://snl.no/AIS

Kleppe, B. (2015, 27. oktober 2017). AIS-satellittar. Hentet 12/1 2021 fra https://www.kystverket.no/Maritime-tjenester/Meldings--og-informasjonstjenester/AIS/AISSat-1-og-AISSat-2/

Kontopoulos, I., Chatzikokolakis, K., Zissis, D., Tserpes, K. & Spiliopoulos, G. (2020). Real-time maritime anomaly detection: detecting intentional AIS switch-off. *International Journal of Big Data Intelligence, 7*(2), 85-96.

Kroodsma, D., Miller, N. & Roan, A. (2017). The Global View of Transshipment: Revised Preliminary Findings. Hentet fra https://globalfishingwatch.org/wp-content/uploads/GlobalViewOfTransshipment_Aug2017.pdf

Kystverket. (2020a, 24. mars 2020). AIS om bord i skip. Hentet 29.10.2020 fra https://www.kystverket.no/Maritime-tjenester/Meldings--og-informasjonstjenester/AIS/AIS-om-bord-i-skip/

Kystverket. (2020b, 2. desember 2020). Brukertilgang AIS Norge. Hentet 27. mars 2021 fra https://www.kystverket.no/Maritime-tjenester/Meldings--og-informasjonstjenester/AIS/Brukartilgang-til-AIS-Norge/

Kystverket. (2021). Ny norsk satelitt fanger opp radarsignaler. Hentet 14. mai 2021 fra https://www.kystverket.no/Nyheter/2021/april/ny-norsk-satellitt-fanger-opp-radarsignaler/

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note, 6*(70), 1.

Last, P., Bahlke, C., Hering-Bertram, M. & Linsen, L. (2014). Comprehensive analysis of automatic identification system (AIS) data in regard to vessel movement prediction. *The Journal of Navigation, 67*(5), 791-809.

Lim, K. (2016). Big data and strategic intelligence. *INTELLIGENCE AND NATIONAL SECURITY, 31*(4), 619-635.

Mandel, D. R. (2020). The occasional maverick of analytic tradecraft. *INTELLIGENCE AND NATIONAL SECURITY, 35*(3), 438-443. Hentet fra https://www.tandfonline.com/doi/pdf/10.1080/02684527.2020.1723830

MarineTraffic. (2018). What is the significance of the AIS Shiptype number? Hentet 18. januar 2021 fra https://help.marinetraffic.com/hc/en-us/articles/205579997-What-is-the-significance-of-the-AIS-Shiptype-number-

MarineTraffic. (2019). *Improving Maritime Situational Awareness Through Big Data Analytics, Machine Learning and Artificial Intelligence*.

Marrin, S. (2002). Homeland security and the analysis of foreign intelligence. *Markle Foundation Task Force on National Security in the Information Age, 15*, 609-637.

Meteorologisk institutt. (2021). Ekstremværet Frank er over. Hentet 2. januar 2021 fra https://www.met.no/nyhetsarkiv/ekstremvaeret-frank-ekstremt-kraftige-vindkast-i-deler-av-nordland-og-troms

Nærings- og fiskeridepartementet. (2018). Om ulovlig, urapportert og uregulert (UUU) fiske. Hentet 15. januar 2021 fra https://www.regjeringen.no/no/tema/mat-fiske-og-landbruk/fiskeri-og-havbruk/1/fiskeri/ulovlig-fiske/om-ulovlig-urapportert-og-uregulert-uuu-fiske/id2579076/

Narkhede, N. (2017). Introducing KSQL: Streaming SQL for Apache Kafka. Hentet 15. januar 2021 fra https://www.confluent.io/blog/ksql-streaming-sql-for-apache-kafka/
NLOD. (2020). Norsk lisens for offentlige data Hentet fra https://data.norge.no/nlod/no/
Noll, M. (2020). Streams and Tables in Apache Kafka: A Primer. Hentet 27. mars 2021 fra https://www.confluent.io/blog/kafka-streams-tables-part-1-event-streaming/
Oksnes, H. H. (2014). *En verdivurdering av Havfisk ASA*.
Omand, D. (2010). *Securing the State* Hurst.
Omand, D. (2014). The future of intelligence: What are the threats, the challenges and the opportunities. *The future of intelligence: Challenges in the 21st century*.
Omand, D. (2019). Et historisk tilbakeblikk. I S. Stenslie, L. Haugom & B. H. Vaage (Red.), *Etterretningsanalyse i den digitale tid - en innføring* (s. 32-50). Bergen: Fagbokforlaget.
Pedersen, M., Alvik, T., Digre, H., Lie, P. W., Bech, H. I., Karlsen, K. M., … Tetmo, H. P. (2019). *NOU 2019: 19 - Framtidens fiskerikontroll* (Norges offentlige utredninger).
Popper, K. R. (2002). *Conjectures and Refutations: The Growth of Scientific Knowledge* Routledge.
Primor, O. (2020, 29.10). 'Going dark' is so 2019. Hentet fra https://wnwd.com/insights/going-dark-is-so-2019/
Schnelle, S. (2018). *Kartlegging av maritime hybride trusler. Kan bruk av stordata og sosial nettverksanalyse bidra til økt maritim situasjonsbevissthet?* Forsvarets høgskole.
Skauen, A. N. (2016). Quantifying the tracking capability of space-based AIS systems. *Advances in Space Research, 57*(2), 527-542.
Skjelland, E., Glærum, S., Beadle, A. W., Endregard, M., Guttelvik, M. S., Hennum, A. C., … Olsen, K. E. (2019). *Hvordan styrke forsvaret av Norge?-Et innspill til ny langtidsplan (2021–2024)* Forsvarets Forskningsinstitutt.
Speier, C., Valacich, J. S. & Vessey, I. (1999). The influence of task interruption on individual decision making: An information overload perspective. *Decision sciences, 30*(2), 337-360.
Statistisk sentralbyrå. (2019). 07842: Registrerte fiskefartøy med motor, etter statistikkvariabel og år. I. Hentet fra https://www.ssb.no/statbank/table/07842/tableViewLayout1/
Stenslie, S. (2019). Å bruke strukturerte analyseteknikker. I S. Stenslie, L. Haugom & B. H. Vaage (Red.), *Etterretningsanalyse i den digitale tid - en innføring* (1. utgave. utg., s. 65-83). Bergen: Fagbokforlaget.
Stenslie, S., Haugom, L. & Vaage, B. H. (2019). Innledning. I S. Stenslie, L. Haugom & B. H. Vaage (Red.), *Etterretningsanalyse i den digitale tid - en innføring* (1. utgave. utg., s. 19-31). Bergen: Fagbokforlaget.
Stølen, K. (2019). *Teknologivitenskap : forskningsmetode for teknologer*. Oslo: Universitetsforlaget.
Stolpe, A., Hansen, B. J. & Halvorsen, J. (2019). *Stordatasystemer og deres egenskaper* (FFI-rapport 18/01676). Forsvarets Forskningsinstitutt. Hentet fra https://www.ffi.no/en/publications-archive/stordatasystemer-og-deres-egenskaper
Stolpe, A., Hansen, B. J., Halvorsen, J. & Opland, E. J. (2020). Experimenting with a big data infrastructure for multimodal stream processing. Hentet fra https://www.ffi.no/en/publications-archive/experimenting-with-a-big-data-infrastructure-for-multimodal-stream-processing
Størdal, J.-M. (2019). Vi må utvikle kultur og kompetanse for et høyteknologisk forsvar. Hentet fra https://www.dn.no/innlegg/forsvarets-forskningsinstitutt/forsvaret/det-teknologiske-skiftet/vi-ma-utvikle-kultur-og-kompetanse-for-et-hoyteknologisk-forsvar/2-1-712839
Svartdal, F. (2019). Bekreftelsestendens. I *Store Norske Leksikon*. Hentet 25. april 2021 fra https://snl.no/bekreftelsestendens
Tu, E., Zhang, G., Rachmawati, L., Rajabally, E. & Huang, G.-B. (2017). Exploiting AIS data for intelligent maritime navigation: A comprehensive survey from data to methodology. *IEEE Transactions on Intelligent Transportation Systems, 19*(5), 1559-1582.

U.S. Coast Guard Navigation Center. (2016). HOW AIS WORKS. Hentet 11. januar 2021 fra https://www.navcen.uscg.gov/?pageName=AISworks
UK Ministry of Defence. (2013). Quick wins for busy analysts. *London: Defence Intelligence, UK*.
Vivento & Kaupang. (2015). *Kartlegging og vurdering av stordata i offentlig sektor*. Kommunal-og moderniseringsdepartementet.
Warner, M. (2008). Intelligence as risk shifting. I *Intelligence Theory* (s. 30-46). Routledge.
Wikipedia. (2017). JSON. Hentet 27. mars 2021 fra https://no.wikipedia.org/wiki/JSON
Wikipedia. (2020). CSV. Hentet 27. mars 2021 fra https://no.wikipedia.org/wiki/CSV
Windward. (2014). Data on the High Seas: An Analysis of the Magnitude and Implications of Growing Data Manipulation at Sea. Hentet fra https://www.arbitrage-maritime.org/fr/Gazette/G36complement/Windward.pdf

Vedlegg A prosjektgodkjennelse fra NSD

NSD NORSK SENTER FOR FORSKNINGSDATA

NSD sin vurdering

Prosjekttittel

Etterretningsanalyse og stordata

Referansenummer

339871

Registrert

11.01.2021 av Lars Roar Uggerud - luggerud@gmail.com

Behandlingsansvarlig institusjon

Forsvarets Høgskole / Forsvarets stabsskole

Prosjektansvarlig (vitenskapelig ansatt/veileder eller stipendiat)

Mass Soldal Lund, maslund@mil.no, tlf: 93025631

Type prosjekt

Studentprosjekt, masterstudium

Kontaktinformasjon, student

Lars Roar Uggerud Dugstad, luggerud@gmail.com, tlf: 92622421

Prosjektperiode

01.08.2020 - 01.06.2021

Status

21.04.2021 - Vurdert

Vurdering (1)

21.04.2021 - Vurdert

Det er vår vurdering at behandlingen vil være i samsvar med personvernlovgivningen så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet 21.04.2021 med vedlegg, samt i meldingsdialogen mellom innmelder og NSD.

FORMÅL
Prosjektet er en masteroppgave ved forsvarets stabsskole som skal undersøke hvordan stordata-analyse

kan berike en spesifikk analyseteknikk innenfor etterretningsfaget (Analyse av konkurrerende hypoteser). Prosjektet vil ta utgangspunkt i undersøkelser om ulovlig, urapportert og uregulert fiske i norske farvann. Fokuset i oppgaven vil ikke være på enkeltskip, men på hvordan stordata-analyse kan berike nevnte analyseteknikk.

BESKRIVELSE
Datagrunnlaget til prosjektet er til kystverkets åpne AIS data fra alle fartøy innenfor et dekningsområde som omfatter norsk økonomisk sone og vernesonene ved Svalbard og Jan Mayen, men med unntak av fiskefartøy under 15 meter og fritidsfartøy under 45 meter. I løpet av en tidsperiode på 24-48 timer vil prosjektet følge de de ulike fartøyenes bevegelser, for se om noen av skipene slår ut på indikatorene for økt sannsynlighet for ulovlig fiske. Studenten har utviklet algoritmer som i samtid søker etter «mørke tid» (Manipulering av posisjonsdata eller at AIS slås av), fartøyenes historikk med UUU og omlasting til sjøs.

TYPE OPPLYSNINGER OG VARIGHET
Prosjektet vil behandle alminnelige personopplysninger og personopplysninger om straffedommer og lovovertredelser frem til 01.06.2021

LOVLIG GRUNNLAG
Prosjektet vil behandle overnevnte kategorier av personopplysninger med grunnlag i at oppgaven er nødvendig for å utføre en oppgave i allmennhetens interesse og for formål knyttet til vitenskapelig forskning.

Lovlig grunnlag for behandlingen av alminnelige personopplysninger er dermed at den er nødvendig for å utføre en oppgave i allmennhetens interesse, jf. personvernforordningen art. 6 nr. 1 bokstav e, samt for formål knyttet til vitenskapelig forskning, jf. personopplysningsloven § 8, jf. personvernforordningen art. 6 nr. 3.

Lovlig grunnlag for behandlingen av personopplysninger om straffedommer og lovovertredelser er at den er nødvendig for formål knyttet til vitenskapelig forskning, jf. personvernforordningen art. 10, jf. art. 9 nr. 2 bokstav j, jf. personopplysningsloven § 9, jf. § 11 første ledd.
Behandlingen er omfattet av nødvendige garantier for å sikre den registrertes rettigheter og friheter, jf. personvernforordningen art. 89 nr. 1.

PERSONVERNPRINSIPPER
NSD vurderer at den planlagte behandlingen av personopplysninger vil følge prinsippene i personvernforordningen:
- om lovlighet, rettferdighet og åpenhet (art. 5.1 a),
- formålsbegrensning (art. 5.1 b), ved at personopplysninger samles inn for spesifikke, uttrykkelig angitte og berettigede formål, og ikke viderebehandles til nye uforenlige formål
- dataminimering (art. 5.1 c), ved at det kun behandles opplysninger som er adekvate, relevante og nødvendige for formålet med prosjektet
- lagringsbegrensning (art. 5.1 e), ved at personopplysningene ikke lagres lengre enn nødvendig for å oppfylle formålet

DE REGISTRERTES RETTIGHETER
Så lenge de registrerte kan identifiseres i datamaterialet vil de ha følgende rettigheter: innsyn (art. 15), retting (art. 16), sletting (art. 17), begrensning (art. 18) og protest (art. 21).

NSD vurderer at det kan unntas fra informasjon jf. personvernforordningen art. 14 nr. 5 b) med begrunnelsen at det vil kreve uforholdsmessig stor innsats å gi informasjon. I vår vurdering har vi vektlagt at det ikke direkte vil registreres personopplysninger, men at det i enkelte tilfeller vil være mulig å identifisere en eier av skip ved å bruke eksterne kilder. Eventuelle skip som har personnavn vil anonymiseres før publikasjon. Behandlingstiden er kort og få personer vil ha tilgang til dataene.

Vi minner om at hvis en registrert tar kontakt om sine rettigheter, har behandlingsansvarlig institusjon plikt til å svare innen en måned.

FØLG DIN INSTITUSJONS RETNINGSLINJER
NSD legger til grunn at behandlingen oppfyller kravene i personvernforordningen om riktighet (art. 5.1 d), integritet og konfidensialitet (art. 5.1. f) og sikkerhet (art. 32).
For å forsikre dere om at kravene oppfylles, må prosjektansvarlig følge interne retningslinjer/rådføre dere med behandlingsansvarlig institusjon.

MELD VESENTLIGE ENDRINGER
Dersom det skjer vesentlige endringer i behandlingen av personopplysninger, kan det være nødvendig å melde dette til NSD ved å oppdatere meldeskjemaet. Før du melder inn en endring, oppfordrer vi deg til å lese om hvilke type endringer det er nødvendig å melde: https://www.nsd.no/personverntjenester/fylle-ut-meldeskjema-for-personopplysninger/melde-endringer-i-meldeskjema Du må vente på svar fra NSD før endringen gjennomføres.

OPPFØLGING AV PROSJEKTET
NSD vil følge opp ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet.

Kontaktperson hos NSD: Kajsa Amundsen

Lykke til med prosjektet!