

NHH



NORGES HANDELSHØYSKOLE

Bergen, Vår 2022

# Anvendelse av kunstig intelligens i revisors vurdering av fortsatt drift

*En eksperimentell studie*

**Thomas Endresen Aase & Mads Matre Lynås**

**Veileder: Jonas Gaudernack**

Masteroppgave, regnskap og revisjon

NORGES HANDELSHØYSKOLE

Dette selvstendige arbeidet er gjennomført som ledd i masterstudiet i regnskap og revisjon ved Norges Handelshøyskole og godkjent som sådan. Godkjenningen innebærer ikke at Høyskolen eller sensorer inntår for de metoder som er anvendt, resultater som er fremkommet eller konklusjoner som er trukket i arbeidet.

## Sammendrag

Revisor sin vurdering av en enhets evne til fortsatt drift er fundamental for opprettholdelsen av tillit i kapitalmarkedene. Vurderingen er imidlertid svært sammensatt, som historisk sett har medført utfordringer i praksis. For formodentlig å øke kvaliteten på denne beslutningen, vil bruk av kunstig intelligens-baserte beslutningsstøtteverktøy kunne være formålstjenlig. Tilgangen på slike verktøy reiser en del praktiske spørsmål, hvilket kan være grunnen til den begrensede bruken i praksis. Det er likevel mye som tyder på at bransjen etterhvert vil bli mottagelig for denne type verktøy. I den anledning ser vi her nærmere på en eventuell påvirkning anvendelsen vil ha, og har utledet følgende forskningsspørsmål: *Påvirker bruk av et KI-basert beslutningsstøtteverktøy vurderingen av fortsatt drift, og reduseres algoritme-aversjon ved kunnskap om algoritmen?*

Den stadige og eksponentielle teknologiske utviklingen, har endret omfanget av den finansielle rapporteringen. Dette har medført utvikling også i revisjonsbransjen, herunder nye arbeidsmetoder og fremgangsmåter for revisor. Tilgang på kunstig intelligens som evner å utøve skjønn, har både fordeler og ulemper for praksis. Dagens regulering hindrer ikke bruk av kunstig intelligens, men tilsynsmyndigheter har uttrykt skepsis. I tillegg er mennesker preget av kognitive «bias», hvilket påvirker effekten av beslutningsstøtteverktøy.

Vårt eksperiment omfattet 41 revisjonsstudenter eller praktiserende revisorer. Respondentene ble tilfeldig allokert til én av tre grupper, med ulik manipulasjon gjennom operasjonalisering av konseptuelle uavhengige variabler. Resultatene fra studien tilsier at bruk av et KI-basert beslutningsstøtteverktøy ikke påvirker vurderingen av fortsatt drift. Vi finner imidlertid at tilliten til algoritmen øker når det gis en forklaring av den, men ikke tilstrekkelig til å redusere algoritme-aversjonen slik den her måles.

## Forord

Denne utredningen er skrevet som en del av masterstudiet i regnskap og revisjon ved Norges Handelshøyskole. Masteroppgaven utgjør 30 studiepoeng og ble skrevet våren 2022.

Arbeidet har blant annet gitt oss mulighet til fordypning innen et fagområde av interesse, samt innsikt i en spennende forskningsverden. Nye arbeidsmetoder, fremgangsmåter og verktøy har medført utfordringer underveis, hvilket har resultert i verdifull læring.

To hovedfagområder preger utredningens innhold, fortsatt drift og kunstig intelligens. Førstnevnte er et stadig aktuelt tema innen vårt fagfelt, og følgelig svært aktuell kunnskap å ta med seg inn i arbeidslivet. Sistnevnte har vært et fremtredende tema i løpet av de siste årene, med betydelig teknologisk suksess. Ettersom teknologien nå har høyt momentum inn i revisjonsbransjen, finner vi forskning på bruken høyst verdifullt.

Med begrenset forskningserfaring, er det nyttig med noen tilbakemeldinger underveis. Vi ønsker i den anledning å rette en stor takk til vår veileder, Jonas Gaudernack, som regelmessig har gitt oss gode råd og innspill. En takk rettes også til medstudenter som deltok i pilottestingen, deltakerne som tok seg tid til å besvare spørreundersøkelsen samt Saad Bin Anis og Ulf Mohrmann for deres råd i forbindelse med de statistiske analysene.

Bergen, juni 2022

---

Thomas Endresen Aase

---

Mads Matre Lynås

---

# Innholdsfortegnelse

<b>SAMMENDRAG</b> .....	<b>II</b>
<b>FORORD</b> .....	<b>III</b>
<b>INNHOLDSFORTEGNELSE</b> .....	<b>IV</b>
<b>1. INNLEDNING</b> .....	<b>1</b>
1.1    MOTIVASJON FOR OPPGAVEN.....	1
1.2    FORSKNINGSSPØRSMÅL.....	1
1.3    FORSKNINGSOMRÅDE.....	2
1.3.1    « <i>Judgment and decision making</i> ».....	2
<b>2. TEORI OG TIDLIGERE FORSKNING</b> .....	<b>4</b>
2.1    PROFESJONELT SKJØNN.....	4
2.1.1 <i>Skjønn</i> .....	4
2.1.2 <i>Profesjonelt skjønn</i> .....	5
2.2    FORTSATT DRIFT.....	6
2.2.1 <i>Aktuell regnskapsregulering</i> .....	6
2.2.2 <i>Sannsynlighetsvurdering</i> .....	7
2.2.3 <i>ISA 570 Fortsatt drift</i> .....	8
2.3    KUNSTIG INTELLIGENS.....	11
2.3.1 <i>Digitalisering av revisjonsbransjen</i> .....	11
2.3.2 <i>Kunstig intelligens</i> .....	12
2.3.3 <i>KI-basert beslutningsstøtteverktøy</i> .....	14
2.3.4 <i>Utfordringer knyttet til KI</i> .....	15
2.4    KI BRUKT I REVISJON.....	19
2.4.1 <i>Revisor versus KI i fortsatt drift-vurderingen</i> .....	21
<b>3. KONSEPTUELL MODELL OG HYPOTESER</b> .....	<b>22</b>
3.1    KONSEPTUELL MODELL.....	22
3.1.1 <i>Konseptuelt rammeverk</i> .....	22
3.1.2 <i>Avhengig variabel</i> .....	23
3.1.3 <i>Uavhengige variabler</i> .....	23

---

3.1.4	<i>Kontrollvariabler</i> .....	24
3.2	HYPOTESEUTFORMING.....	24
3.2.1	<i>Hypotese 1: Implementering av KI i vurderingen</i> .....	25
3.2.2	<i>Hypotese 2: Effekten av forklaring på vurderingen av saksforholdet</i> .....	26
3.2.3	<i>Hypotese 3: Algoritme-aversjon</i> .....	27
<b>4.</b>	<b>METODE</b> .....	<b>28</b>
4.1	FORSKNINGSDESIGN .....	28
4.1.1	<i>Forskningstilnærming</i> .....	28
4.1.2	<i>Metodiske valg</i> .....	29
4.1.3	<i>Forskningsstrategi</i> .....	30
4.1.4	<i>Faktorielt design</i> .....	30
4.1.5	<i>Between-/within subjects</i> .....	32
4.2	DATAINNSAMLING .....	33
4.2.1	<i>Utvalg</i> .....	33
4.2.2	<i>Utforming av eksperimentell undersøkelse</i> .....	34
4.2.3	<i>Pilottest</i> .....	35
4.2.4	<i>Praktisk gjennomføring</i> .....	36
4.3	ANALYSEMETODE.....	36
4.3.1	<i>Dataanalyse</i> .....	36
4.3.2	<i>Forutsetninger for analysen</i> .....	37
4.4	FORSKNINGSETIKK OG PERSONVERN .....	38
4.5	RELIABILITET OG VALIDITET.....	38
<b>5.</b>	<b>RESULTATER</b> .....	<b>41</b>
5.1	RENSING AV DATA .....	41
5.1.1	<i>Manipulasjonssjekk</i> .....	42
5.2	DESKRIPTIV STATISTIKK .....	42
5.3	HYPOTESETESTING .....	43
5.3.1	<i>Forutsetninger analyse</i> .....	44

---

5.3.2	<i>Resultater hypotese 1: Implementering av KI i vurderingen</i> .....	44
5.3.3	<i>Resultater hypotese 2: Effekten av forklaring på vurderingen av saksforholdet</i> .....	45
5.3.4	<i>Resultater hypotese 3: Algoritme-aversjon</i> .....	47
<b>6.</b>	<b>DISKUSJON OG KONKLUSJON</b> .....	<b>49</b>
6.1	DISKUSJON AV HYPOTESER .....	49
6.1.1	<i>Hypotese 1</i> .....	49
6.1.2	<i>Hypotese 2</i> .....	50
6.1.3	<i>Hypotese 3</i> .....	51
6.2	IMPLIKASJONER FOR PRAKSIS .....	53
6.3	KONKLUSJON .....	53
6.4	BEGRENSNINGER .....	54
6.5	FORSLAG TIL VIDERE FORSKNING .....	55
	<b>LITTERATURLISTE</b> .....	<b>56</b>
	<b>VEDLEGG</b> .....	<b>67</b>
	VEDLEGG 1: EKSPERIMENTELT CASE OG UNDERSØKELSE .....	67
	VEDLEGG 2: CRONBACH'S ALPHA .....	71
	VEDLEGG 3: SHAPIRO-WILK TEST .....	71
	VEDLEGG 4: HYPOTESE 1 .....	72
	VEDLEGG 5: HYPOTESE 2 .....	73
	VEDLEGG 6: HYPOTESE 3 .....	75

---

## Figuroversikt

Figur 1: Libby-bokser.....	22
Figur 2: 2x2 faktorielt design .....	31
Figur 3: Gjennomsnittssvar for gruppe 1 og 4 – profesjonelt skjønn.....	44
Figur 4: Gjennomsnittssvar for gruppe 2 og 4 – profesjonelt skjønn.....	46
Figur 5: Gjennomsnittssvar for gruppe 2 og 4 – algoritme-aversjon .....	47

## Tabelloversikt

Tabell 1: Deskriptiv statistikk .....	43
Tabell 2: Deskriptiv statistikk og p-verdier – hypotese 1.....	45
Tabell 3: Deskriptiv statistikk og p-verdier – hypotese 2.....	46
Tabell 4: Deskriptiv statistikk og p-verdier – hypotese 3.....	48

# 1. Innledning

## 1.1 Motivasjon for oppgaven

Revisor har en viktig rolle i finansmarkedene, og omtales i revisorloven (2020, § 9-1) som «allmennhetens tillitsperson» ved utførelse av lovfestet revisjon. Med det følger et ansvar hvis mangelfulle etterlevelse kan få fatale konsekvenser, hvilket Arthur Andersen er et eksempel på (Store norske leksikon, 2020). Som følge av revisjonsbransjens økonomiske attraktivitet, har konkurransen tilspisset seg over flere år. Dette har medført et press på marginene og følgelig et behov for kontinuerlig effektivisering. Her som i andre bransjer, fører det til knapphet på tid og ressurser.

Begrensninger i tid og ressurser for å opprettholde lønnsomheten, gjør revisor sårbar for feil. Det har derfor blitt tatt i bruk en rekke teknologiske verktøy frem til i dag, som gjør revisor i stand til å kunne fokusere på de viktigste, mest risikofylte og skjønnsmessige områdene. Den teknologiske utviklingen muliggjør i dag også maskinell skjønnsutøvelse, og det er nettopp her utgangspunktet for vårt forskningsprosjekt gjør seg gjeldende. Innføringen i bruk av digitale verktøy gjennom kurset «Digital revisjon» ved NHH, samt dagsaktualiteten av teknologisk skjønnsutøvelse, gjorde at vi fattet interesse for bruk av digitale verktøy i revisjonen.

## 1.2 Forskningsspørsmål

Sarbanes Oxley Act (SOX) kan på mange måter sies å være et symbol på finansskandalene rundt forrige tusenårsskifte. Formålet med implementeringen av loven var å skjerpe den finansielle rapporteringen, for å unngå at lignende skandaler som Enron og WorldCom skulle gjentas (Store norske leksikon, 2018, 2019b). Tillit til regnskapene er i aller høyeste grad fortsatt avgjørende for å opprettholde stabilitet i finansmarkedene. Derfor er det også at tilsynsorganer som finanstilsynet, følger nøye med ved endringer i regnskapsbransjen. Tilbakeholdenheten til finanstilsynet smitter naturligvis over til bransjen, og har antageligvis begrenset bruken av visse teknologiske verktøy noe.



Hvorvidt regnskaps- og revisjonsstandardene åpner for bruk av kunstig intelligens (KI), i et revisjonsperspektiv, er en drøftelse som vi har valgt å ikke gå nærmere inn på. Konsensus i bransjen er imidlertid at kunstig intelligens ikke kan brukes ukritisk, og er derfor lite utbredt på tidspunktet denne utredningen skrives. Interessant finner vi det derfor å se nærmere på hvorvidt bruken av kunstig intelligens påvirker en konkret beslutnings-/vurderingssituasjon. Forskningsspørsmålet er utformet som følger:

***Påvirker bruk av et KI-basert beslutningsstøtteverktøy vurderingen av fortsatt drift, og reduseres algoritme-aversjon ved kunnskap om algoritmen?***

En eventuell observert effekt av å ta i bruk et KI-verktøy vil i praksis føre til enten en mer eller mindre kostnadseffektiv revisjon, avhengig av både revisor og KI sine vurderinger. Eksempelvis vil tilfellet der revisor vurderer at det foreligger hendelser eller forhold som kan skape tvil av betydning om enhetens evne til fortsatt drift og KI-verktøyet at fortsatt driftsforutsetningen kan legges til grunn, teoretisk føre til at ytterligere revisjonshandlinger reduseres dersom revisor hensyntar KI-verktøyet.

Som angitt i forskningsspørsmålet ønsker vi også å se om algoritme-aversjonen reduseres ved kunnskap om algoritmen. Dette kan sies å være en sekundærobservasjon. Algoritme-aversjon har åpenbare implikasjoner for implementeringen av KI-verktøy i revisjonen, som gjør det aktuelt å måle effekten av denne i en vurdering av fortsatt drift.

## 1.3 Forskningsområde

### 1.3.1 «Judgment and decision making»

Forskning relatert til «judgment and decision making» (JDM) innen regnskap og revisjon tiltok på 70- og 80-tallet, grunnet behov for beskrivende forskning på området, bredere aksept for eksperimentelle metoder samt forskere sin økte kunnskap om kjernedisipliner som kognitiv psykologi (Ashton & Ashton, 1995, s. 3). Viktigheten av «judgment» i revisjon uttrykkes implisitt gjennom revisjonsstandardene, og gjør at JDM-forskning er et av de meste utbredte forskningsområdene (Trotman, 2001, s. 181). Formålet med JDM-forskning er å beskrive hvordan og i hvilken utstrekning revisor utøver skjønn og fatter beslutninger, samt finne ut hvordan disse kan forbedres (Ashton & Ashton, 1995, s. 139; Trotman, 2001, s. 181).

---

Vår utredning er ment å bidra til JDM-forskningen, ved å se nærmere på hvordan implementeringen av kunstig intelligens påvirker skjønnsutøvelsen i forbindelse med en vurdering av fortsatt drift-forutsetningen.

## 2. Teori og tidligere forskning

I dette kapitlet ser vi nærmere på et utvalg relevant teori og tidligere forskning. Teorien er strukturert fra det generelle til det spesifikke, hvor sistnevnte her omhandler revisjonsteori. Det foreligger en rekke tidligere empiri på områdene vi ser nærmere på, og en avgrensning vil derfor være nødvendig.

### 2.1 Profesjonelt skjønn

#### 2.1.1 Skjønn

Filosofen Ronald Dworkin definerer skjønn som et rom for frihet innenfor rammen av visse restriksjoner fastlagt av en autoritet (Dworkin, 1977, i Fossheim & Ingierd, 2015, s. 103). «Skjønnsutøvelse handler altså ikke om å ha frihet til å gjøre hva en vil, men om å ha frihet til å tolke og anvende det rammeverket som autoriteten har fastsatt» (Fossheim & Ingierd, 2015, s. 103). Lorentzen (2005) definerer skjønnsutøvelse med «situasjoner der beslutningstaker må ta stilling til hvilke kriterier som er relevante i den konkrete beslutningssituasjonen, og hvordan de ulike kriteriene skal forstås og anvendes». Skjønndefinisjonen er ikke entydig, og vil ha noe ulik forståelse i forskjellige kontekster. Generelt sett kan man imidlertid forstå utøvelse av skjønn til forskjell fra mer regelbaserte beslutninger, som «viser til situasjoner der beslutningstakers fullmakter er mer bundet gjennom formelle normer som legger føringer på beslutningstakers vurderinger og beslutninger» (Lorentzen, 2005).

Utøvelsen av skjønn kan deles i tre komponenter, (1) «å ta stilling til hvilke hensyn som skal telle med i den aktuelle beslutningssituasjonen», (2) hensynene skal operasjonaliseres, «det vil si at de skal gis et presist innhold» og (3) det må «utøves skjønn når man skal ta stilling til hvordan hensynene skal avveies mot hverandre (Lorentzen, 2005). Parallelt med denne prosessen vil det være nyttig å ha kjennskap til potensielle fallgruver. Goodin (1986, i Fossheim & Ingierd, 2015, s. 102) peker på fire fallgruver ved utstrakt bruk av skjønn i moderne velferdsstater: (1) Maktutnyttelse, (2) vilkårlighet, (3) uforutsigbarhet og (4) illegitim innblanding. Sistnevnte er ikke relevant å diskutere i vårt tilfelle, og utelates fra videre drøftelse.

Maktutnyttelse handler ifølge Fossheim og Ingierd (2015, s. 102) i denne sammenheng om «mulighet til å utøve sosial og moralsk kontroll over den som trenger bestemte ytelser eller stønader». En potensiell konsekvens vil være at det utvikler seg et mistillitsforhold med negative samfunnseffekter, som følge av at unødvendige ressurser går med til å rette opp i det skjeve maktforholdet. Med vilkårlighet menes at den som utøver fullstendig skjønn, til forskjell fra regelbaserte avgjørelser, ikke har plikt til å begrunne avgjørelsen, hvor mangel på begrunnelsesplikt fører til vilkårlige avgjørelser (Goodin, 1986, i Fossheim & Ingierd, 2015, s. 102). «Den tredje fallgraven er at skjønnsbaserte avgjørelser er uforutsigbare og derfor fører til usikkerhet og mangel på trygghet» (Fossheim & Ingierd, 2015, s. 102).

### **2.1.2 Profesjonelt skjønn**

Profesjonelt skjønn kan sies å være et underbegrep av skjønn. Begrepet innebærer å benytte og kombinere erfaring, fleksibilitet, informerte vurderinger og stadig selvstyring i en kvalifisert avgjørelse (Tripp, 1993). Distinksjonen mot mer generelle diskresjonære beslutninger er at utøvelse av profesjonelt skjønn krever mer spesifikke bakgrunnskunnskaper. Eksempelvis peker Jacob et al. (2021, s. 4226) på den avgjørende betydningen av profesjonell skjønnsutøvelse i en sykepleier-logistikk-kontekst, hvor spesielt fagkunnskap og praktisk erfaring spiller en sentral rolle. Blant de mange andre fagområdene hvor profesjonelt skjønn står sentralt, finner man revisjon.

Profesjonelt skjønn defineres i ISA<sup>1</sup> 200 som «anvendelsen av relevant opplæring, kunnskap og erfaring innenfor rammen av standarder for revisjon, regnskap og etikk når det fattes informerte beslutninger om hvilke handlingsplaner som er hensiktsmessige etter omstendighetene ved revisjonsoppdraget», og beskrives som avgjørende for et vellykket revisjonsoppdrag (IAASB, 2009a, pkt 13; Puthukulam et al., 2021, s. 1184). Eksempler på praktisk anvendelse av profesjonelt skjønn er i forbindelse med valg av type og omfang av ytterligere revisjonshandlinger, herunder innsamling av revisjonsbevis og kvaliteten av disse. Et annet eksempel er vurderingen av om noe er «vesentlig», som blant annet gjør seg gjeldende i forbindelse med å vurdere usikkerheten til en enhets evne til fortsatt drift.

---

<sup>1</sup> International Standards on Auditing

## 2.2 Fortsatt drift

Fortsatt drift er stadig et svært aktuelt tema i regnskap- og revisjonsbransjen. Blant annet presiserte ESMA<sup>2</sup>, i en offentlig uttalelse like etter pandemierklæringen av Covid-19, viktigheten av å gi informasjon om vurderingen av fortsatt drift i forbindelse med den halvårlige finansielle rapporteringen (ESMA, 2020). Finanstilsynet sluttet seg til uttalelsen like etter (Finanstilsynet, 2020).

Senere støtteordninger (Regjeringen, u.å.), utvikling i sentrale pengemarkedsrenter, større usikkerhet rundt framtidsutsikter og bransjeendringer, har aktualisert vurderingen ytterligere. I tillegg fører stadige endringer i satser, avgifter, regler og frister knyttet til skatter og avgifter (Skatteetaten, u.å.), i enkelte tilfeller til at selskaper som ellers ikke ville vært økonomisk levedyktige, får forlenget levetid. Dette også som følge av at det gjentatte ganger har blitt innført tiltak for å motvirke de økonomiske konsekvensene av pandemien, når det har vist seg å være nødvendig.

Fortsatt drift-forutsetningen innebærer at en virksomhet forventes å fortsette i overskuelig fremtid (Rødssæteren & Gøbel, 2010, s. 14). Vurderingen av forutsetningen er først og fremst bemyndiget ledelsen i det aktuelle selskapet, men revisor har også oppgaver og plikter vedrørende forholdet.

### 2.2.1 Aktuell regnskapsregulering

Regnskapspliktige<sup>3</sup> skal utarbeide årsregnskapet under forutsetning om fortsatt drift så lenge det ikke er sannsynlig at virksomheten vil bli avviklet (regnskapsloven, 1998, § 4-5). Forutsetningen om fortsatt drift er et av regnskapslovens ti grunnleggende regnskapsprinsipp, angitt i lovens kapittel 4, hvor et brudd på forutsetningen potensielt påvirker innregning, måling, presentasjon og noteopplysninger (Bernhoft et al., 2018, s. 118).

---

<sup>2</sup> European Securities and Markets Authority

<sup>3</sup> Jf. regnskapsloven, 1998, § 1-2.

---

Som følge av EØS-avtalen er Norge underlagt EUs lovgivning om utarbeidelse av finansiell informasjon i tråd med internasjonale regnskapsstandarder (IFRS<sup>4</sup>). Av IAS<sup>5</sup> 1 følger det at «finansregnskapet skal utarbeides under forutsetning om fortsatt drift, med mindre ledelsen enten har til hensikt å avvike foretaket eller legge ned virksomheten, eller ikke har noe realistisk alternativ til dette» (IASB, 2005, pkt. 25). Ordlyden er noe annerledes enn i regnskapsloven, som angitt over. Lien (2012) argumenterer imidlertid for at en avveining av rettskildefaktorene tilsier at regnskapsloven § 4-5 har samme innhold som IAS 1. Dette er i tråd med det pågående arbeidet om internasjonal harmonisering av regnskapslovgivningen (Norsk Regnskapsstiftelse, 2021).

I ny revisorlov av 2020 tydeliggjøres viktigheten av vurderingen knyttet til fortsatt drift. Lovens (2020) § 9-4 fjerde ledd angir at revisor «skal særlig utvise profesjonell skepsis<sup>6</sup> ved vurdering av regnskapsestimater knyttet til virkelige verdier, nedskrivninger av eiendeler, avsetninger, transaksjoner med nærstående parter og fremtidige kontantstrømmer av betydning for foretakets evne til fortsatt drift». Den utførlige angivelsen av fortsatt drift i regnskapslovgivningen og revisjonsstandarder, medfører en bevisstgjøring for revisor som formodentlig resulterer i veloverveide vurderinger.

## 2.2.2 Sannsynlighetsvurdering

Hva som menes med «sannsynlig» i regnskapslovens § 4-5 er ikke eksplisitt angitt av lovgiver. Sannsynlighetskravet ble imidlertid diskutert blant annet av regnskapslovutvalgets innstilling til ny regnskapslov (NOU 1995:30; Pettersen, 2009, s. 22), samt av finansdepartementet i proposisjon til Stortinget (Prop. 42 (1997-1998) pkt. 6.7.4), i forbindelse med ny regnskapslov. Verken NOUen eller proposisjonen tydeliggjorde i særdeleshet hvordan kravet skulle forstås.

I en kommentarutgave til regnskapsloven uttrykkes «sannsynlig» som over 60-70 prosent (Johnsen & Kvaal, 1998, s. 188). Dette forstås å være i tråd med proposisjon til Stortinget (Prop. 42 (1997-1998), pkt. 6.7.5), som setter terskelen «noe høyere» enn 50 prosent.

---

<sup>4</sup> International Financial Reporting Standards

<sup>5</sup> International Accounting Standards

<sup>6</sup> «En holdning som innebærer at revisor stiller spørsmål og er oppmerksom på forhold som kan indikere mulig feilinformasjon som følge av feil eller misligheter, og foretar en kritisk vurdering av revisjonsbevis» (IAASB, 2009a, pkt. 13).

Kvantifisering av sannsynlighetsbegreper innen god regnskapsskikk finner man også angitt i NRS 13 *Usikre forpliktelser og betingede eiendeler* (Norsk Regnskapsstiftelse, 2003, pkt. 3.2), som angir et sannsynlighetsintervall mellom 50 og 70 prosent for «sannsynlighetsovervekt». For øvrig presiserer standarden at usikkerhet i regnskapsmessige problemstillinger som oftest ikke vil kunne tallfestes med presise sannsynlighetsprosenter (Norsk Regnskapsstiftelse, 2003, pkt. 3.2).

Ifølge Johnsen og Kvaal må tidshorizonten for vurderingen defineres, om det skal gi mening å tallfeste sannsynlighetsnivået (Lien, 2012, s. 24). Den aktuelle tidshorizonten ved vurdering av fortsatt drift ble omtalt i lovforarbeidene til regnskapsloven, hvor finansdepartementet uttrykte at den økonomiske levetiden til anleggsmidlene skulle legges til grunn (Prop. 42 (1997-1998), pkt. 11.5). I en lovkommentar til regnskapsloven finner Handeland et al. (2020, s. 207) det i så fall naturlig å dele opp perioden i en nær periode (typisk frem til neste balansedag) og en lang periode (tiden deretter), hvor den nære tillegges mest vekt.

Tatt i betraktning den foreliggende uklarheten av hva som menes med «sannsynlig», samt vanskeligheten av å kvantifisere denne i praksis, gjør det noe underlig at lovgiver (til tross for flere muligheter) har unnlatt å klargjøre innholdet i bestemmelsen (Lien, 2012, s. 25). Vurderingen vil i enkelte tilfeller være tvetydig som følger, hvilket kan føre til uenigheter når også revisor skal gjøre seg opp en vurdering av saksforholdet.

### **2.2.3 ISA 570 Fortsatt drift**

En direkte konsekvens av at forutsetningen om fortsatt drift er av grunnleggende betydning for regnskapet, er at revisors oppgaver og plikter vedrørende forholdet utførlig er omtalt i en egen revisjonsstandard: ISA 570. Standarden tilsier at revisor skal vurdere hvorvidt regnskapsprodusent har grunnlag for å utarbeide regnskapet under forutsetning om fortsatt drift, og om det foreligger en «vesentlig usikkerhet<sup>7</sup>» til enhetens evne til fortsatt drift (IAASB 2016a, pkt. 6).

Av ISA 570 fremgår det at når «fortsatt drift legges til grunn for regnskapet, utarbeides regnskapet under den forutsetning at enheten er i drift og vil fortsette sin virksomhet i

---

<sup>7</sup> For definisjon, se ISA 570 (IAASB, 2016a, pkt. 18).

---

overskuelig fremtid» (IAASB, 2016a, pkt 2). Hva standardsetter legger i overskuelig fremtid er ikke videre definert i standarden, men må forstås som en avveining av alle tilgjengelig fakta og omstendigheter knyttet til enhetens evne til å fortsette virksomheten, som foreligger på det til enhver tid gjeldende vurderingstidspunktet. Lovutvalget til ny regnskapslov presiserte at foretakets evne til fortsatt drift må forstås mer konkret enn at selskapet fortsetter sin eksistens, ved at også samme virksomhet skal fortsette (Prop. 42 (1997-1998) pkt. 6.7.3).

Tidshorizonten er som nevnt sentral i vurderingen for om fortsatt drift-forutsetningen kan legges til grunn. Johnsen og Kvaal (1998, s. 188) illustrerer dette ved at sannsynligheten for at et foretak skal være avviklet i morgen eller i neste uke er neglisjerbar, men betydelig sett i et femtiårs perspektiv. Det internasjonale standardiseringsorganet IAASB<sup>8</sup> har derfor gitt presiseringer angående den aktuelle tidshorizonten. Revisor skal vurdere samme periode som ledelsen har lagt til grunn i sin vurdering, begrenset nedad til tolv måneder fra balansedagen (IAASB, 2016a, pkt. 13). I tillegg krever standarden at revisor skal forespørre ledelsen om indikasjoner på usikkerhet til fortsatt drift, «ut over perioden som ledelsen har lagt til grunn i sin vurdering (IAASB, 2016a, pkt 15). Dette må generelt sett forstås å være ut over nær periode, som drøftet over.

Standarden tilsier at revisor skal «vurdere hvorvidt det foreligger hendelser eller forhold som kan skape tvil av betydning om enhetens evne til fortsatt drift» (IAASB, 2016a, pkt. 10). Det poengteres i det påfølgende punktet i standarden, at oppmerksomheten rundt eventuelle indikasjoner skal opprettholdes under hele revisjonen. I tillegg kreves det en evaluering av ledelsens vurdering av enhetens evne til fortsatt drift (IAASB, 2016a, pkt. 12).

Under vurderingen som foretas, er det argumentert for at forutsetningen om fortsatt drift skal betraktes som en nullhypotese (NOU 1995: 30). Med det menes at man ved fravær av indikasjoner, eller mangel på overbevisende observasjoner, på usikkerhet til enhetens evne til fortsatt drift, skal ta utgangspunkt i at forutsetningen om fortsatt drift kan legges til grunn. I innstilling til Stortinget (NOU 1995: 30) fremgår det at styrken på indikasjonene på at nullhypotesen skal forkastes, bestemmes av en avveining mellom type I og type II feil. Type I feil innebærer å feilaktig forkaste en sann nullhypotese, og type II feil at man unnlater å forkaste nullhypotesen når den skulle vært forkastet (Braut, 2021). Type I feil kan føre til en

---

<sup>8</sup> International Auditing and Assurance Standards board



selvoppfyllende profeti, altså at konkurssannsynligheten øker ved uttrykt usikkerhet rundt fortsatt drift (Kjellevoll, 2019, s. 31). Type II feil kan på sin side føre til at markedet ikke informeres om problematikken i tide, og kan i tillegg føre til erstatningsansvar for revisor (Revisorloven, 2020, § 11-1).

Dersom revisor kommer frem til at det foreligger hendelser eller forhold som kan skape tvil av betydning om enhetens evne til fortsatt drift, skal det utføres ytterligere revisjonshandlinger (IAASB, 2016a, pkt. 16). Standarden angir konkret hvilke handlinger som skal utføres, blant annet skal sannsynligheten og gjennomførbarheten av ledelsens planer vurderes, herunder hvorvidt de prognostiserte fremtidige kontantstrømmene (der dette foreligger og er relevant) er pålitelige og med substans. Dette for å fastslå om det foreligger en vesentlig usikkerhet (IAASB, 2016a, pkt. 16).

Revisor må dernest konkludere på om forutsetningen kan legges til grunn, og være oppmerksom på eventuelle konsekvenser for revisjonsberetningen. Konsekvensene beror på ledelsens vurderinger, revisors konklusjoner, opplysninger i regnskapet og gjeldende omstendigheter ved virksomheten (IAASB, 2016a, pkt. 21-23). Utfallet vil være enten umodifisert konklusjon, forbehold eller negativ konklusjon i samsvar med krav i ISA 705 (IAASB, 2016b). I tilfeller der forutsetningen om fortsatt drift kan legges til grunn, det foreligger en vesentlig usikkerhet og det er gitt tilstrekkelige opplysninger, skal det legges til et eget avsnitt i beretningen med overskrift «Vesentlig usikkerhet knyttet til fortsatt drift» (IAASB, 2016a, pkt. 22).

Revisors vurdering av enhetens evne til fortsatt drift får konsekvenser både for den aktuelle enheten, men også for dens interessenter. I verste fall kan en feilaktig forkastelse av nullhypotesen om fortsatt drift (type I-feil), føre til et unødvendig tap av arbeidsplasser og verdiskaping. Som følge av at vurderingen er diskresjonær av natur, vil kognitive bias og eksterne faktorer kunne påvirke vurderingen (Colson, 2019). Dette er, blant andre, noe som har aktualisert bruk av kunstig intelligens i beslutningsprosessen.

---

## 2.3 Kunstig intelligens

### 2.3.1 Digitalisering av revisjonsbransjen

Revisjonsbransjen har de siste årene gjennomgått, og står fortsatt i, en digital omstilling (Fjørtoft, 2018, s. 26; Kinserdal, 2017). Som følge av den teknologiske fremgangen har de store revisjonsselskapene investert tungt i teknologi som skal brukes i revisjonsarbeidet, omtalt Audit Data Analytics-verktøy (KPMG, 2018; Deloitte, 2021; EY, 2021; PwC, 2020). AICPA (2017, egen oversettelse) definerer ADA som «kunsten å oppdage og analysere mønstre, identifisere anomalier, og hente ut annen nyttig informasjon i data som ligger til grunn for eller relatert til formålet ved revisjonsoppdraget gjennom analyser, modellering og visualisering med hensikt å planlegge eller utføre revisjonen». Bruken av ADA må sees i sammenheng med stordata.

Stordata defineres som datasett som er for store eller komplekse å håndtere ved hjelp av tradisjonelle metoder og teknologier for databehandling (PwC, u.å.a). Det som skiller stordata fra tradisjonell data er at informasjonen karakteriseres med høyt volum, høy hastighet og stor variasjon. Med stort volum menes mengde data, hastighet referer til at dataen er tilgjengelig i sanntid og variasjon at dataen stammer fra flere forskjellige kilder, eksempelvis GPS-systemer, video og lydfiler (Yoon et al., 2015, s. 432). Når data stammer fra flere forskjellige kilder, vil den være både strukturert og ustrukturert. Strukturert data kan formateres i tabeller, som for eksempel transaksjonsdata. Ustrukturert data er det som ikke kan struktureres, eksempelvis bilder og video (PwC, u.å.a). For å analysere den ustrukturerte dataen i revisjonssammenheng, vil ADA være nyttig.

Til tross for store investeringer i ADA-verktøy, er selve bruken av ADA lav blant revisjonsselskapene. Selskapene begrunner dette med usikkerheten til om regulatoriske myndigheter godtar revisjonsbevis generert fra ADA (Eilifsen, 2020, s. 76). I dag benyttes ADA-verktøy i hovedsak til å gi supplerende revisjonsbevis, istedenfor å erstatte eksisterende metoder. Det medfører at revisjonsprosessen blir mindre effektiv, da man i realiteten gjør den samme arbeidsoppgaven flere ganger (Eilifsen, 2020, s. 76-77). Selv om revisjonsstandardene ikke forbyr bruken av ADA som revisjonsbevis, oppfordrer de heller ikke til det. I henhold til ISA 530 må revisor, ved hjelp av profesjonelt skjønn, vurdere hvilke handlinger som resulterer i hensiktsmessige og tilstrekkelige revisjonsbevis for å redusere revisjonsrisikoen til et

akseptabelt nivå (IAASB, 2009b). Finanstilsynet har informert revisjonsselskapene om at de først vil inspisere en revisjon gjennomført med ADA-verktøy, og deretter evaluere om verktøyene er egnet til formålet. Det er dermed opp til revisjonsselskapene å bevise at digitale revisjonsmetodikker er minst like sikre som tradisjonelle revisjonsmetoder. Dette gjør at revisjonsselskapene vegrer seg for å ta verktøyene i bruk, i frykt for søksmål eller sanksjoner fra tilsynsmyndighetene (Eilifsen, 2020, s. 80; Kinserdal, 2017). Ettersom ADA er et svært bredt begrep, har vi valgt å avgrense oppgaven til å se på kunstig intelligens.

### **2.3.2 Kunstig intelligens**

Det finnes mange definisjoner på hva kunstig intelligens er. Hovedessensen i definisjonene er at det er en teori og utvikling av datasystemer i stand til å gjennomføre oppgaver som vanligvis krever menneskelig intelligens (Burgess, 2018, s. 5). Begrepet kunstig intelligens ble først introdusert i 1955 av John McCarthy, og siden den gang har optimismen rundt hva KI kan gjøre vært stor. Det er imidlertid først i nyere tid at kunstig intelligens virkelig har fått fotfeste. Dette skyldes i hovedsak den teknologiske fremgangen i lagringsmulighetene og prosesseringskraften til datamaskinene, samt den enorme tilgangen på data (Brynjolfsson & McAfee, 2019, s. 7; Kokina & Davenport, 2017, s. 116).

Burgess (2018) mener at kunstig intelligens har tre målsetninger: (1) innhente informasjon, (2) bestemme hva som skjer og (3) forstå det som skjer. Videre forsøker han å operasjonalisere begrepet kunstig intelligens i åtte hovedområder: bildegjenkjenning, stemmegjenkjenning, søk, klyngeanalyse, språkbehandling, optimalisering, prediksjon og forståelse. De fire første operasjonaliseringene handler om å bruke dataen, eksempelvis ved å finne mønstre og sammenhenger i store ustrukturerte datasett. De tre neste operasjonaliseringene handler om å bruke dataen, og den siste om å forstå den. Kunstig intelligens er imidlertid per dags dato mindre utviklet til å forstå det som skjer.

Fagterminologien skiller mellom smal og sterk kunstig intelligens (Burgess, 2018, s. 4; PwC, u.å.b). I dag finnes det flere eksempler der KI brukes på spesifikke oppgaver, eksempelvis Tesla som utvikler selvkjørende biler, og Facebook sine ansiktsgjenkjenningsalgoritmer. Dette er kjennetegn på smal KI, også kalt første generasjon KI (Kaplan & Haenlein, 2019, s. 16). KI-applikasjonen fungerer svært effektivt til det den er ment for, men klarer ikke løse oppgaver eller problemer utover det den er trent til. Målet er å utvikle en maskin som kan løse problemer den i utgangspunktet aldri var designet for å løse. Det er dette som kalles for sterk, eller

---

generell kunstig intelligens (Burgess, 2018, s. 4; Kaplan & Haenlein, 2019, s. 16). I dag er det i hovedsak smal kunstig intelligens som benyttes i praksis.

## *Maskinlæring*

Maskinlæring er et grunnleggende begrep innenfor kunstig intelligens, og defineres som en maskin eller et system sin evne til å lære, uten å spesifikt bli programmert til det (Samuel, 1952, s. 211). Fremgangsmåten ved maskinlæring er at man gir maskinen definerte algoritmer, som ved bruk av statistiske metoder lærer maskinen å finne mønstre i datasettet (Tidemann & Elster, 2022). Algoritmen blir trent opp ved hjelp av et treningsdatasett, og det kreves store mengder data for at denne skal fungere effektivt og nøyaktig. Ut fra treningsdataen vil algoritmen utvikle en modell som skal være i stand til å identifisere sammenhenger i data som ligner treningsdataen, og basert på dette gir modellen et estimert resultat (Datatilsynet, 2018, s. 6).

Det skilles i hovedsak mellom to former for maskinlæring, veiledet læring og ikke-veiledet læring. Forskjellen mellom de to formene ligger i informasjonen man gir algoritmen. Med veiledet læring gir man algoritmen forhåndskategoriserte data, og basert på disse er målet at den skal lære seg forskjellen på kategoriene. Eksempelvis gir man algoritmen et stort datasett av ulike dyr, hvor man på forhånd har kategorisert hvilke dyr som er hva. Ut fra dette vil algoritmen lære seg forskjellen på dyrene. Det er denne læringsmetoden som har gitt mest suksess de siste årene (Brynjolfsson & McAfee, 2019, s. 12).

Ved ikke-veiledet læring vil man gi data uten forhåndskategorier, og maskinen skal da selv finne mønstre i datasettet basert på likheter og ulikheter mellom dataen. Følger vi eksempelet med dyr, vil det bety at algoritmen kun mottar et datasett med alle dyrene, men ingen forhåndsdefinert informasjon om disse. Det blir følgelig opp til algoritmen selv å kategorisere dyrene. For begge læringsmetodene er det nødvendig med omfattende treningsdatasett, gjerne med flere millioner eksempler, for at algoritmene skal kunne predikere nøyaktige resultater (Brynjolfsson & McAfee, 2019, s. 11; Burgess, 2018, s. 20; Datatilsynet, 2018, s. 7-8).

Tidligere har det i større grad vært fokus på å finne algoritmen som gir best resultater, fremfor å se på dataen (Russel & Norvig, 2016, s. 27). For ikke-veiledet læring har flaskehalsen vært tilgang på data. Studier viser at ikke-veiledet læringsalgoritmer presterer bedre enn veilede i tekstanalyse, når treningsdatasettet er stort nok (Banko & Brill, 2001, s. 32). I dagens teknologiske verden med stordata, vil tilgangen på data være enklere enn tidligere.

Algoritmen som har stått bak mye av suksessen med maskinlæring er dyp læring. Dette er en underkategori av maskinlæring, som baserer seg på kunstige nevrane nettverk (Brynjolfsson & McAfee, 2019, s. 11). Strukturen i algoritmen er en forenklet modell av den menneskelige hjernen, med nevroner som kommuniserer med hverandre. Det kunstige nevrane nettverket består normalt av tre hovedlag: et input-lag, et eller flere skjulte lag og et output-lag. Kompleksiteten til algoritmen øker når antall skjulte lag øker. Nettverket er bygd opp av nevroner, der hvert nevron er koblet mot et annet nevron med ulik styrke. Hvert nevron får informasjon (input), og benytter statistiske metoder for å generere informasjon til et nytt nevron (output). Informasjonen går fra et nevron til et annet avhengig av styrken mellom nevronene. Dersom nevronet finner at informasjonen overstiger den kritiske verdien, sendes informasjonen videre og det tilkoblede nevronet aktiveres. (Hardesty, 2017; IBM, 2020; Tidemann, u.å.).

Med dyp læring vil nevronene, ved hjelp av store mengder treningsdata, selv komme frem til hva som er viktig for å skille dataen fra hverandre. Algoritmen gjør i tillegg egne tilpasninger for å øke nøyaktigheten i prediksjonene (IBM, 2020).

### **2.3.3 KI-basert beslutningsstøtteverktøy**

Inntoget av kunstig intelligens skaper store muligheter. Det vil gi en effektivitetsgevinst ved at maskiner kan gjennomføre tidkrevende, rutinemessige oppgaver på en brøkdel av tiden av det mennesker klarer (Issa et al., 2016, s. 9). I tillegg vil det føre til høyere beslutningskvalitet ved at algoritmene ved bruk av statistiske metoder har en høyere treffsikkerhet i flertallet av tilfellene, sammenlignet med menneskelige eksperter (Elkins et al., 2013, s. 251). Dette har gjort at flere bransjer har tatt i bruk KI-baserte beslutningsstøtteverktøy for å bedre nøyaktigheten og effektiviteten i beslutningsprosessen (Moxander & Axente, 2021, s. 1).

Det er i hovedsak to faktorer som gir algoritmen forutsetning til å ta mer nøyaktige beslutninger enn mennesker: (1) mindre «bias», også kalt systematisk skjevhet og (2) mindre tilfeldig variasjon (Kahneman et al., 2016, s. 40-41). Algoritmene tar kun høyde for informasjonen som er relevant for beslutningen, uavhengig av hvor mye irrelevant informasjon som er inkludert i datasettet. Den vil dermed opptre objektivt i større grad enn mennesker, som kan være påvirket av blant annet personlige preferanser, dagsform og i hvilken rekkefølge man mottar informasjon, selv i situasjoner de skal opptre objektivt. I tillegg vil en algoritme i de aller fleste tilfeller komme frem til den samme beslutningen, gitt at den

---

mottar lik informasjon. En studie viste at leger som ble fremstilt det samme røntgenbilde på to forskjellige tidspunkter, endret sin vurdering i 20 % av tilfellene (Kahneman et al., 2011, i Sjøstad, 2019).

Når man benytter KI-baserte beslutningsstøtteverktøy vil mennesker ta skjønnsmessige beslutninger basert på egen ekspertise og vurderinger, samt ta hensyn til anbefalingene til beslutningsstøtteverktøyet. For at en person skal ta hensyn til verktøyet, er det avgjørende at vedkommende har tillit til systemet (Vereschak et al., 2021, s. 1).

Tillit er et komplekst begrep, og det finnes flere ulike definisjoner. Lee og See (2004, s. 54, egen oversettelse) definerer tillit som «en holdning som utøves av en agent for å hjelpe et individs mål i en situasjon karakterisert av usikkerhet og sårbarhet». Likhetstrekkene i de fleste definisjonene er villigheten til å være sårbar, og forventninger om atferd hos tillitsmottaker. Disse må være til stede for at et tillitsforhold skal kunne eksistere, hvor uttrykt atferd er forankret i de nevnte faktorene.

Sårbarhet er den første nøkkelfaktoren som er nødvendig for tillit. Med sårbarhet vil det være en usikkerhet knyttet til utfallet. En syk pasient vil være sårbar ettersom det vil være usikkerhet knyttet til behandlingen fra doktoren (Vereschak et al., 2021, s. 10). I tilfeller uten sårbarhet, og dermed ingen usikkerhet, vil det ikke være nødvendig med tillit (Rousseau et al., 1998, s. 395). Videre er det nødvendig at man har en positiv forventning hos tillitsmottaker. Mer konkret må man ha en forventning om at motparten vil handle på en bestemt måte, og at negative utfall ved å stole på anbefalingen er usannsynlig (Vereschak et al., 2021, s. 10).

### **2.3.4 Utfordringer knyttet til KI**

Selv om det er store gevinster og muligheter ved å ta i bruk kunstig intelligens som beslutningsstøtteverktøy, viser tidligere studier at mennesker vegrer seg for å ta i bruk disse systemene. Vi ser her nærmere på problemene knyttet til teknologiens transparens, og algoritme-aversjon.

#### ***“Black box” teoremet***

Utviklingen innenfor maskinlæring har gjort at man har beveget seg bort fra regelbaserte maskinlæringsmetoder som ekspertsystemer, til mer avanserte dyplæringsmetoder med kunstig nevralt nettverk. Dette gjør at mennesker i mindre grad har innsikt i hvordan

algoritmen arbeider, og kompleksiteten i de dype nevrane nettverkene med flere millioner interaksjoner gjør det tilnærmet umulig å etterprøve hvordan algoritmen kommer frem til sin beslutning. Problematikken rundt åpenheten til algoritmen refereres ofte til som «black box»-teoremet (Teknologirådet, 2018, s. 52; Zednik; 2019, s. 4). Lite innsikt i beslutningen til algoritmen kan skape flere utfordringer, spesielt ettersom det i tiden fremover er sannsynlig at kunstig intelligens vil ta over flere arbeidsoppgaver som vil påvirke mennesker i det dagligdagse.

Transparensproblematikken rundt kunstig intelligens benyttet som beslutningsverktøy, gjør det svært vanskelig å oppdage feilaktige eller umoralske beslutninger. Dette gjør at man vil være eksponert mot uriktige beslutninger, eller at algoritmen er trent opp på et skjevt datamateriale. Konsekvensen er at algoritmen systematisk kan ta feil beslutning i enkelte situasjoner (Deloitte, 2019). Dette var tilfellet i Chicago, der algoritmer trent opp på historisk data ble benyttet til å predikere hvilke personer som mest sannsynlig ville være involvert i skyteepisoder. Datagrunnlaget bar preg av minoritetsdiskriminering, og programmet ble tatt ut av bruk i 2019 (Gorner & Sweeney, 2020; Hao 2019; Rieland, 2018).

«Black box»-problematikken har i tillegg vist seg gjeldende i situasjoner der utvikleren har mulighet til å forklare beslutningen til algoritmen. Problemet viste seg gjeldende i USA, der en algoritme basert på kunstig intelligens ble brukt til å bistå i beslutningen av fengselsstraff. Da den tiltalte ønsket innsyn i vurderingen til algoritmen, som var svært delaktig for den idømte fengselsstraffen, ble han nektet innsikt fordi utvikler ikke ville risikere å røpe forretningshemmeligheter (Israni, 2017).

### *Algoritme-aversjon*

Til tross for at algoritmene gjentatte ganger har vist seg å utkonkurrere menneskelige eksperter, er det en generell motstand mot å ta i bruk kunstig intelligens i beslutningssituasjoner (Castelo et al., 2019, s. 809; Dietvorst et al., 2015, s. 114). Dette fenomenet blir ofte referert til som algoritme-aversjon. Innenfor studier knyttet til algoritme-aversjon, har det utarbeidet seg to ulike forståelser av begrepet. Dietvorst et al. (2015, s. 114) finner i sin studie at mennesker utvikler algoritme-aversjon etter at man har observert at algoritmen har feilet. Derimot finner Castelo et al. (2019, s. 809) at mennesker i noen tilfeller har en generell aversjon mot algoritmen, selv før man kjenner til algoritmen sin

---

prediksjonsevne. Vi har videre i utredningen valgt å avgrense til å se på den generelle algoritme-aversjonen.

Jussupow et al. (2020, s. 4-5) har avdekket flere målemetoder for å måle algoritme-aversjon. I den første metoden måles algoritme-aversjon ut fra hvordan man vektlegger vurdering eller hjelp fra en algoritme kontra et menneske. Den andre målemetoden tar utgangspunkt i at man først kommer opp med et eget estimat, for deretter å bli eksponert for en vurdering fra både en menneskelig ekspert og en algoritme. Felles for målemetodene er at dersom man vektlegger menneskets vurdering mer enn algoritmen, vil algoritme-aversjon være til stede.

### **Forklaringer til algoritme-aversjon**

Sjåstad (2019) forklarer algoritme-aversjon ut fra fire psykologiske faktorer. Den første faktoren er at mennesker har en overdreven tro på egne ferdigheter, som fører til at man tror man er bedre til å predikere enn det man i realiteten er. To forhold må være til stede for å ha mulighet til å utarbeide reell menneskelig ekspertise: (1) oppgaven man jobber med må være innenfor relativt stabile omstendigheter, og (2) at man mottar umiddelbar tilbakemelding på vurderingen (Kahneman & Klein, 2009, s. 520). I realiteten er disse forholdene sjeldent til stede samtidig, og man vil derfor ofte overvurdere egen prediksjonsevne (Sjåstad, 2019).

Den andre faktoren har direkte sammenheng med forståelsen til Dietvorst et al. (2015), altså at man mister tillit til algoritmen når man ser den feile. Dette gjør man selv om man observerer at mennesker gjør feil oftere enn algoritmen. En feil gjort av en algoritme vil redusere tilliten mer enn en feil gjort av en menneskelig ekspert. Det er altså ikke symmetri i hvordan man vurderer feil av algoritmer og mennesker (Dietvorst et al., 2015, s. 115). Denne atferden kan forklares ut fra to synspunkter. Mennesker har en forventning om at algoritmen er feilfri, i motsetning til mennesker som man vet er imperfekte. Når man da observerer algoritmen feile, vil dette komme som en overraskelse sammenlignet med når mennesker feiler, og tilliten til algoritmen vil reduseres mer (Madhavan & Wiegmann, 2007, s. 284). Atferden kan også forklares ut fra at mennesker har et ønske om perfektjon. Selv der man vet at algoritmen har en høyere prediksjonsevne, vil man stole på mennesket fordi man i større grad forventer at disse kan gi en feilfri prediksjon (Einhorn, 1986, s. 390; Highhouse, 2008, s. 340). Dette kan ses i sammenheng med at man har overdreven tro på egne ferdigheter. I tillegg tar mennesker ofte utgangspunkt i at en observert feil av algoritmen vil være en systematisk feil. I motsetning til dette forventer man at mennesker i større grad evner å lære av sine feil, selv om også algoritmer nå evner å lære av sine feil (Raman et al., 2012, s. 1934)



Den tredje forklaringen er at mennesker er sosiale vesen, og at prediksjonsevnen til algoritmen ikke utelukkende vil være det eneste som påvirker valget mellom maskin og menneske. I noen situasjoner vil den menneskelige faktoren være viktig, og enkelte tjenester er dermed mindre egnet for automatisering enn andre. Den siste faktoren skyldes frykten for tapt individualitet. Styrken til algoritmen er at den fokuserer på nøkkelfaktorer, og filtrerer bort irrelevant informasjon. Utfordringen med dette er at mennesket kan oppfatte algoritmen som for generell, ved at den ikke tar hensyn til spesifikke særtrekk ved egen situasjon. Effekten er særlig gjeldende innenfor helsetjenesten, og spesielt i tilfeller der pasienten anser sin egen situasjon som spesielt unik (Longoni et al., 2019, s. 643). I slike situasjoner bør algoritmen heller brukes som et beslutningsstøtteverktøy, fremfor et substitutt for menneskelig beslutning, da algoritme-aversjonen kan reduseres når man får inntrykk av at beslutningsprosessen ikke er helautomatisk (Sjåstad, 2019).

### **Tiltak for å redusere algoritme-aversjon**

Algoritme-aversjon er et problem for implementeringen av kunstig intelligens, og kan føre til store konsekvenser og kostnader. Det kan dermed svare seg å se på tiltak for å redusere aversjonen, ved å øke tilliten til algoritmen. For å styrke tilliten til en algoritme er det viktig at brukeren forstår de teknologiske mulighetene (Chao et al., 2016, s. 122-123). Det er i den sammenheng viktig å gi informasjon om de iboende begrensningene til algoritmen, slik at man selv kan velge hvilke tilfeller man skal stole på vurderingen til beslutningsstøtteverktøyet (Zhang et al., 2020, s. 295). Ved å gi informasjon om algoritmens begrensninger knyttet til treffsikkerhet i prediksjonsevnen, vil man sørge for at menneskers forventninger til algoritmen samsvarer med dens faktiske prestasjonsevne. Det vil da ikke komme som en overraskelse om algoritmen feiler.

Yin et al. (2019, s. 6) fant at informasjon om treffsikkerheten til en algoritme har stor effekt på tilliten til den. Studien viser at tilliten øker desto høyere treffsikkerhet algoritmen har. Dette er sammenfallende med Castelo et al. (2019, s. 821) som i en lignende studie fant at algoritme-aversjonen avtok når man fikk informasjon om algoritmens overlegenhet ovenfor mennesker. Dette var spesielt gjeldende for objektive oppgaver, men funnene var også signifikante for mer subjektive oppgaver (Castelo et al., 2019, s. 817).

Flere studier viser at aversjonen avtar når en selv får muligheten til å gjøre justeringer i algoritmen (Dietvorst et al., 2018, s. 1167). Ved å få være med på å bestemme hvordan algoritmen skal ta sine beslutninger, vil altså tilliten til den øke. Et problem med menneskelig

---

interaksjon i algoritmen i beslutningssammenhenger, er at det ofte reduserer treffsikkerheten til algoritmen (Lim & O'Connor, 1995, s. 162). Det vil derimot være begrenset hvor store endringer man kan gjøre. Dermed vil algoritmen fortsatt ha verdi, enn om man kun skulle brukt menneskelige eksperter (Dietvorst et al., 2018, s. 1167).

Det er også gjennomført studier på hvordan man kan skape emosjonell tillit til algoritmer. Emosjonell tillit er definert som en atferd som primært motiveres av en positiv tilknytning til objektet (Lewis & Weigert, 1985, s. 972, egen oversettelse). Tilknytning til selskapet som utvikler algoritmen kan ha påvirkning på tilliten til selve algoritmen (Hengstler et al., 2016, s. 106). Dette følger av grunnleggende teori om at man vil ha høyere tillit til en mer pålitelig og troverdig kilde, som i dette tilfellet er algoritmeutvikler (Sternthal et al., 1978, s. 252).

## 2.4 KI brukt i revisjon

De fire store revisjonsselskapene rapporterer om tunge investeringer i teknologiske verktøy basert på kunstig intelligens. KPMG har inngått en teknologiallianse med IBM for å bruke kunstig intelligens på utarbeidelse av regnskapet (KPMG, 2019). PwC vant i 2017 Audit Innovation of the Year for deres analyseverktøy GL.Ai, som analyserer alle transaksjoner og kontoer for å finne anomalier (PwC, 2017). Deloitte har i samarbeid med Kia utarbeidet analyseverktøyet Argus som benytter maskinlæring til å analysere, identifisere og hente ut informasjon fra elektroniske dokumenter (Deloitte, 2017).

En rekke tidligere studier viser at revisjonsbransjen kan effektiviseres ved å ta i bruk kunstig intelligens (Kokina & Davenport, 2017, s. 116). Tidligere har fokuset vært rettet mot å implementere kunstig intelligens for å automatisere arbeidsintensive og repetitive arbeidsoppgaver (Agnew 2016; Rapoport, 2016). Abdolmohammadi (1999, s. 68) gjorde en analyse av revisjonsprosessen, og fant at substanstesting inneholdt de mest strukturerte arbeidsprosessene. I senere tid har det vært mer fokus på å bruke kunstig intelligens for å innhente data, ved at analyseverktøyene innhenter relevant informasjon fra ulike kilder, slik at revisor kan benytte tiden på områder der det utøves profesjonelt skjønn (Agnew 2016; Brennan et al., 2017). Som følge av dette vil kvaliteten på revisjonen øke, ved at revisor blant annet får en bedre forståelse for selskapet og dets omgivelser.

Selv om kunstig intelligens har, og en videre implementering vil gi, store positive gevinster, er det flere etiske dilemmaer som medfører problemer for kunstig intelligens i revisjonsbransjen. Dette gjelder særlig ettersom den teknologiske utviklingen gjør at kunstig intelligente verktøy går fra å være assisterende ved å gjøre arbeidsintensive, rutinepregede arbeidsoppgaver, til å benyttes som beslutningsstøtteverktøy (Munoko et al., 2020, s. 211). Et av de største problemene knyttet til kunstig intelligens i revisjon er «black box»-problematikken. Revisor skal som kjent utøve profesjonelt skjønn i forbindelse med vurderingen av påliteligheten og tilstrekkeligheten til revisjonsbeviset. Dersom man ikke kan forklare beslutningsprosessen til revisjonsbeviset, vil man ha problemer med å vurdere påliteligheten (Munoko, 2020, s. 219). I tillegg kan det oppstå problemer om hvorvidt det er revisor eller algoritmen som skal holdes ansvarlig.

Ved gjennomførelse av en revisjon samler man ofte inn revisjonsbevis fra flere kilder, og man får da ofte «konkurrerende revisjonsbevis». Forskning viser at dersom en vurderer påliteligheten til kildene forskjellig, vil man vektlegge den antatt mest pålitelige kilden mer enn den andre i sin vurdering (Birnbbaum, 1976, s. 428; Birnbbaum & Stegner, 1979, s. 69). Algoritme-aversjon kan gjøre at revisor overvurderer beviset mottatt fra menneskelige eksperter eller ledelsen fremfor bevis mottatt fra kunstig intelligens, selv om beviset mottatt fra ledelsen kan være preget av subjektive estimater (Commerford et al., 2021, s. 173). I en revisjonssammenheng står man ofte ovenfor komplekse estimater som innehar høy usikkerhet, eksempelvis ved vurderingen av fortsatt drift forutsetningen (Carson et al., 2013, s. 360; Pettersen, 2009, s. 21). Algoritmer kan i slike tilfeller gjøre flere feil. Dersom det er tilfellet vil aversjonen naturligvis øke, noe som underbygges av tidligere studier på området (Dietvorst & Bharti, 2020, 1302; Castelo et al., 2019, s. 811).

Ettersom revisjonsklientene i større grad tar i bruk stordata i sine virksomheter, og revisjonsselskapene har investert tungt i digitale revisjonsverktøy, har standardsettere utarbeidet ulike arbeidsgrupper for å tilpasse revisjonsstandardene til den digitale utviklingen. Både IAASB og PCAOB<sup>9</sup> har i sine strategiske utviklingsplaner av revisjonsstandardene for perioden 2020-2023 og 2020-2024, bruken av ADA som et av sine hovedfokusområder i forbindelse med utviklingen av revisjonsstandardene (IAASB, 2019; PCAOB, 2019).

---

<sup>9</sup> Public Company Accounting Oversight Board

---

### 2.4.1 Revisor versus KI i fortsatt drift-vurderingen

I forbindelse med vurderingen av fortsatt drift må man vurdere bevis opp mot hverandre, og vektingen av de ulike bevisene kan være avgjørende for totalvurderingen av enhetens evne til fortsatt drift. Det gjør at revisor i stor grad kan påvirkes av systematisk skjevhet og variasjon i vurderingen. Ulike psykologiske faktorer, som rekkefølgen på mottatte bevis (recency bias), kan også påvirke vurderingen selv om revisor skal opptre objektivt.

Slike faktorer er med på å redusere treffsikkerheten til revisor sin vurdering. Tidligere forskning viser at 90 % av selskapene som mottar et avsnitt om fortsatt drift usikkerhet overlever (type I feil), og at omtrent halvparten av selskapene som går konkurs ikke har et avsnitt om usikkerhet til enhetens evne til fortsatt drift i beretningen året før det går konkurs (type II feil) (Carson et al., 2013, s. 356). Den store andelen type I feil skyldes i hovedsak at statistikken vurderer det som en feil dersom selskapet ikke har gått konkurs det påfølgende året. Revisor sin oppgave er å informere om usikkerheten til enhetens evne til fortsatt drift, ikke predikere konkurser. Selv om disse har en viss sammenheng, kan det være langt fra en enhets usikkerhet til fortsatt drift, til faktisk konkurs. For en mer korrekt vurdering av type I feil bør man inkludere andre variabler over en lengre tidsperiode, som gjeldsstrukturering og fusjoner. Nogler (1995, s. 65) tok utgangspunkt i en bredere definisjon av et selskap sin usikre overlevelsessevne. Han vurderte en lengre periode etter utgitt revisjonsberetning med et avsnitt knyttet til usikkerhet til enhetens evne til fortsatt drift, og finner at overlevelseshraten er 35%, ikke 90%.

For å øke nøyaktigheten i en kompleks vurdering kan man benytte KI-baserte beslutningsstøtteverktøy. Studier der KI-verktøy er trent opp på årsregnskaper, ser på forskjellen i de finansielle nøkkeltallene på selskaper som har mottatt/ikke mottatt en presisering om usikkerhet til fortsatt drift i samme bransje. Verktøyet hadde en treffsikkerhet på godt over 90 %, og reduserte type I feil til 3 %, og type II feil til 2 % (Chi & Chu, 2021, s. 15; Chi & Shen, 2022, s. 11; Jan, 2021, s.16). Sammenlignet med revisor ser vi at algoritmen har en vesentlig høyere treffprosent. Svakheten er at algoritmen er trent opp på datasett som kan inneholde feil, ettersom treningsdatasettet består av tidligere årsregnskap med revisjonsberetninger som kan inneholde både type I og II feil. Likevel er resultatene oppløftende, og KI-baserte beslutningsverktøy kan gi et vesentlig bedre beslutningsgrunnlag for revisor samt bidra til bedre vurderinger og lavere feilmarginer.

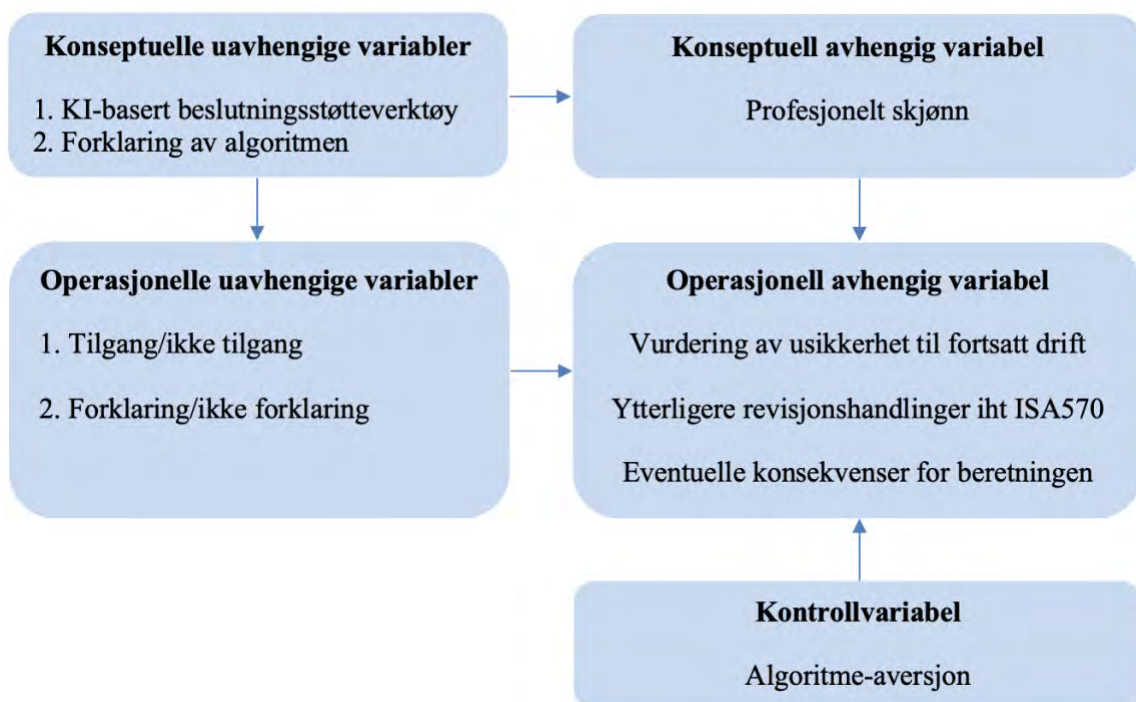
### 3. Konseptuell modell og hypoteser

I kapittelet som følger vil vi se nærmere på den konseptuelle modellen som lagt til grunn for forskningen, forklare og drøfte valg vi har tatt med bakgrunn i relevant teori, samt utforme og begrunne valg av hypoteser.

#### 3.1 Konseptuell modell

##### 3.1.1 Konseptuelt rammeverk

Robert Libby utviklet en konseptuell modell for eksperimentelle design på 70-tallet. Rammeverket beskriver prosessen der «forskeren søker å tilpasse observasjoner av endringer i avhengige- og uavhengige variabler for å sikre at eventuelle endringer i førstnevnte forårsakes av endringer i sistnevnte» (Libby, 1981, s. 11). Formålet med vårt eksperiment er å kartlegge hvorvidt et KI-basert beslutningsstøtteverktøy påvirker vurderingen av fortsatt drift, og hvilken effekt algoritme-aversjon har i denne sammenhengen. Med utgangspunkt i rammeverket som beskrevet, kan forholdet mellom de aktuelle variablene utledes på følgende måte.



Figur 1: Libby-bokser

---

Med bakgrunn i teori og tidligere forskning har vi en presumsjon om at tilgang på et beslutningsstøtteverktøy basert på kunstig intelligens, vil påvirke utøvelsen av profesjonelt skjønn. I tillegg har vi i litteraturgjennomgangen funnet at algoritme-aversjon har en vesentlig betydning for bruken av kunstig intelligens, hvor forklaring påvirker aversjonen. Tilgang på et KI-basert beslutningsstøtteverktøy samt forklaring av det, er følgelig effektene ønsker vi teste gjennom vår forskning, og er visualisert gjennom konseptualiseringen av variablene. Variablene er dernest operasjonalisert for å gjøre de målbare i praksis. Som drøftet ovenfor har den hurtigvoksende tilgangen på «black-box» KI-modeller ført med seg skeptikere blant berørte parter, herunder beslutningstakere og reguleringsmyndigheter. Ettersom respondentene våre er beslutningstakere, ønsket vi derfor å se nærmere på hvorvidt algoritme-aversjon påvirker effekten forklaring har på utøvelsen av profesjonelt skjønn.

### 3.1.2 Avhengig variabel

Avhengig variabel er «en variabel som i noen tilfeller endres som følge av endringer i andre variabler; observert resultat eller resultat fra manipulasjon av en annen variabel» (Saunders et al., 2016, s. 179). Konseptuell avhengig variabel er «profesjonelt skjønn», og vi antar med det at «endringer» i tilgangen på KI-baserte beslutningsstøtteverktøy og/eller forklaring av algoritmen, i noen tilfeller vil påvirke utøvd skjønn.

Den avhengige variabelen er operasjonalisert gjennom tre diskresjonære vurderinger/beslutninger: (1) *Vurdering av usikkerhet til fortsatt drift*, (2) *ytterligere revisjonshandlinger* og (3) *eventuelle konsekvenser for beretningen*. Vi valgte tre operasjonaliseringer av den avhengige variabelen for å øke sannsynligheten for å avdekke en eventuell effekt. De tre nivåene er antatt å være proxy for skjønnsutøvelse, og spørres eksplisitt om i undersøkelsen (vedlegg 1).

### 3.1.3 Uavhengige variabler

En uavhengig variabel er en «variabel som blir manipulert eller endret for å måle påvirkningen på en avhengig variabel» (Saunders et al., 2016, s. 179). De konseptuelle uavhengige variablene er «KI-basert beslutningsstøtteverktøy» og «forklaring av algoritmen». Med KI-basert menes en algoritme som er i stand til å gjennomføre oppgaver som vanligvis krever menneskelig intelligens og beslutningsstøtteverktøy med at algoritmen er ment å supplere egne vurderinger i forbindelse med beslutningen. Ytterligere forklaring av algoritmen er antatt

å øke dens tillit blant respondentene, og følgelig altså påvirke utøvelsen av profesjonelt skjønn i den aktuelle vurderingen.

Den første uavhengige variabelen operasjonaliseres gjennom følgende to tilfeller: (1) *Revisor har ikke tilgang på KI-verktøy som vurderer at regnskapet kan utarbeides under forutsetning om fortsatt drift*, og (2) *Revisor har tilgang på KI-verktøy som vurderer at regnskapet kan utarbeides under forutsetning om fortsatt drift*.

Operasjonaliseringen av den andre uavhengige variabelen var ganske enkelt forklaring eller ikke forklaring. Ved «ikke forklaring» ble det kun gitt informasjon om at et KI-verktøy var tatt i bruk, og konklusjonen. Ved «forklaring» ble det gitt informasjon om blant annet utvikler, treffsikkerhet i prediksjonene, treningsdatasettet og fordeler ved algoritmer kontra mennesker, i tillegg til konklusjonen. Metodiske valg for hvordan de ulike nivåene av de uavhengige variablene testes i praksis, er videre omtalt i kapittel 4.

### 3.1.4 Kontrollvariabler

Kontrollvariabler er «uønskede men målbare variabler som det er nødvendig å holde konstant for å unngå påvirkning på effekten av den uavhengige variabelen på den avhengige» (Saunders et al., 2016, s. 713). Formålet er å eliminere effekten av irrelevante faktorer på resultatene (Spector, 2020). Vi inkluderte som nevnt kontrollvariabelen algoritme-aversjon for å teste hvorvidt aversjon påvirker effekten den uavhengige variabelen «forklaring» har på utøvelsen av profesjonelt skjønn. Tillit til algoritmen samt vektlegging av denne i egen vurdering, antas å måle algoritme-aversjon i undersøkelsen. Dette utgjør følgelig operasjonaliseringen av denne variabelen.

## 3.2 Hypoteseutforming

En hypotese er «en gjetning, antagelse eller forklaring som synes rimelig ut fra foreliggende kunnskap, og som man forsøker å avkrefte eller bekrefte» (Tranøy, 2020). Implisitt i «hypotese» gjør to hypoteser seg gjeldende simultant: alternativ- og nullhypotese (Trochim et al., 2016, s. 17). Med utgangspunkt i teori og tidligere forskning utarbeides en presumsjon som utgjør alternativhypotesen, det er denne man ønsker å gi støtte til (Trochim et al., 2016, s. 17). Øvrige utfall utgjør nullhypotesen, som forkastes når resultatene er signifikante.

---

### 3.2.1 Hypotese 1: Implementering av KI i vurderingen

En vurdering av fortsatt drift-forutsetningen kan være svært kompleks, der informasjonen i stor grad inneholder subjektive estimater, samt at den kan inneha forhold som trekker i begge retninger. Revisor sin utøvelse av profesjonelt skjønn vil dermed være avgjørende for kvaliteten på beslutningen. Tidligere studier viser at bruk av KI som beslutningsstøtteverktøy vil øke kvaliteten på beslutningene (Zhang et al., 2020, s. 295).

En studie av børsnoterte selskaper i USA fra 2001 til 2010 viser at 60 % av selskapene som gikk konkurs, mottok en revisjonsberetning uten en presisering angående usikkerheten knyttet til fortsatt drift (type II feil) (Carson et al., 2013, s. 356). En slik feilvurdering kan som tidligere drøftet, medføre uheldige konsekvenser. Nyere forskning der ulike former for kunstig intelligens benyttes for å predikere fortsatt drift usikkerhet, viser at enkelte modeller kan ha en treffsikkerhet på over 90 % (Chi & Shen, 2022, s. 16).

Med bakgrunn i tidligere forskning er vår antagelse at bruk av et KI-verktøy som konkluderer med at fortsatt drift forutsetningen er til stede, reduserer revisor sin usikkerhet knyttet til enhetens evne til fortsatt drift, sammenlignet med der man ikke har tilgang til et slikt verktøy i beslutningsprosessen. Valget av retning på konklusjonen til verktøyet var førende for en ensidig hypotese. Caset vi har utarbeidet inneholder stor usikkerhet, noe vi fikk bekreftet gjennom pilottestene vi utførte. Usikkerheten gjør at utfallsrommet er større når konklusjonen til verktøyet er at fortsatt drift-forutsetningen kan legges til grunn, hvilket medfører større sannsynlighet for at vi ser en effekt av verktøyet. Med utgangspunkt i dette har vi utformet følgende null- og alternativhypotese:

**$H_0(1)$ :** Revisor vurderer at det foreligger uendret eller mer usikkerhet til enhetens evne til fortsatt drift, ved tilgang på et KI-verktøy som konkluderer med at fortsatt drift-forutsetningen er til stede.

**$H_1(1)$ :** Revisor vurderer at det foreligger mindre usikkerhet til enhetens evne til fortsatt drift, ved tilgang på et KI-verktøy som konkluderer med at fortsatt drift-forutsetningen er til stede.



### 3.2.2 Hypotese 2: Effekten av forklaring på vurderingen av saksforholdet

Algoritmer har selv i komplekse situasjoner med høy usikkerhet, vist seg å ta mer nøyaktige beslutninger enn mennesker. Likevel har mennesker en tendens til å ha en generell aversjon mot å ta i bruk algoritmer (Castelo et al., 2019, s. 809). Tidligere studier viser at informasjon om algoritmens overlegenhet ovenfor mennesker og kunnskap om dens funksjoner og iboende begrensninger, kan redusere algoritme-aversjon (Burton et al., 2020, s. 223; Castelo et al., 2019, s. 817; Zhang et al., 2020, s. 295).

For å stole på et verktøy, er det avgjørende at man har tillit til hvor det kommer fra. Det kan være vanskelig å opparbeide tillit til et produkt som man selv ikke har hatt samhandling med tidligere, og det vil da være viktig at man stoler på produsenten/tilbyderen. Tilbyderen må dermed anses å være pålitelig og levere produkter av høy kvalitet for at en skal ha tillit til dens produkter (Kennedy et al., 2001, s. 75). I tillegg vil kjennskap til algoritmens treffsikkerhet sammenlignet med mennesker, øke tilliten til algoritmen (Castelo et al., 2019, s. 817; Yin et al., 2019, s. 1).

Vår antagelse er at økt informasjon om verktøyet vil øke tilliten til algoritmen. Respondentene mottar informasjon om hvordan algoritmen fungerer, dens nøyaktighet og hvordan den presterer i forhold til revisor i sin vurdering. I tillegg har vi gitt informasjon om at det er Microsoft, en svært anerkjent teknologiprodusent, som har laget beslutningsverktøyet. Informasjonen samlet sett var ment å underbygge påliteligheten til verktøyet, som øker sannsynligheten for at det kan benyttes som et revisjonsbevis. Med utgangspunkt i dette antar vi at revisor, ved mottatt informasjon om verktøyet, blir mindre usikker på enhetens evne til fortsatt drift, når verktøyet konkluderer med at fortsatt drift forutsetningen kan legges til grunn, sammenlignet med når man ikke får informasjon om verktøyet. Dette gir oss følgende hypoteser:

***H<sub>0</sub>(2):** Revisor vurderer at det foreligger uendret eller mer usikkerhet til enhetens evne til fortsatt drift, når det gis en forklaring av KI-verktøyet sammenlignet med når det ikke gis en forklaring.*

***H<sub>1</sub>(2):** Revisor vurderer at det foreligger mindre usikkerhet til enhetens evne til fortsatt drift, når det gis en forklaring av KI-verktøyet sammenlignet med når det ikke gis en forklaring.*

### 3.2.3 Hypotese 3: Algoritme-aversjon

Med bakgrunn i drøftet teori, er vår presumsjon at algoritme-aversjon er faktoren som vil ha størst innvirkning på effekten den uavhengige variabelen «forklaring» har på utøvelsen av profesjonelt skjønn. Med hypotese 3 ønsker vi derfor å teste respondentene sin algoritme-aversjon direkte, for å få en forståelse av hvorvidt en eventuell effekt av forklaring kan tilskrives redusert algoritme-aversjon.

Gjennom å analysere resultatene fra hypotese 3, får vi testet om det foreligger en generell algoritme-aversjon og om den aktuelle forklaringen som gis, vil redusere aversjonen. Tidligere studier viser at særlig informasjon om treffsikkerheten til algoritmen har en innvirkning for den uttrykte tilliten til den (Yin et al., 2019, s. 6). Med utgangspunkt i dette har vi utformet følgende hypoteser.

***H<sub>0</sub>(3):** Revisor sin algoritme-aversjon er uendret eller høyere når det gis en forklaring av KI-verktøyet.*

***H<sub>1</sub>(3):** Revisor sin algoritme-aversjon reduseres når det gis en forklaring av KI-verktøyet.*

## 4. Metode

I kapittel 4 redegjøres det for metoden lagt til grunn for forskningen. Vi ser nærmere på noen sentrale valg i forbindelse med utarbeidelse av metoden, samt drøfter avveininger ved disse.

### 4.1 Forskningsdesign

Forskningsdesign er en overordnet plan for fremgangsmåten lagt til grunn for å svare på forskningsspørsmålet (Saunders et al., 2016, s. 163, egen oversettelse). De argumenterer videre for at forskningsspørsmålet har en sentral rolle i forbindelse med utarbeidelsen av forskningsdesignet. Vi minner derfor her på utformingen av forskningsspørsmålet:

*Påvirker bruk av et KI-basert beslutningsstøtteverktøy vurderingen av fortsatt drift, og reduseres algoritme-aversjon ved kunnskap om algoritmen?*

Forskningsdesignet vårt er et klassisk eksperiment ettersom respondentene tilfeldig tildeles en av tre grupper. Det valgte designet er generelt sett det mest hensiktsmessige ved ønske om å etablere en årsak-virkning-sammenheng (Trochim et al., 2016, s. 221). Vi har i det følgende valgt å se nærmere på fem elementer ved designet.

#### 4.1.1 Forskningstilnærming

Det skilles ofte mellom to hovedtilnærminger til teoriutvikling i en forskningskontekst, deduksjon og induksjon (Saunders et al., 2016, s. 144). «Deduksjon er en utledning eller logisk slutning fra noe allment til et enkelttilfelle» (Store norske leksikon, 2019a). «Induksjon er en vitenskapelig metode som på grunnlag av enkelte, individuelle utsagn eller erfaringer søker å komme frem til mer omfattende generelle utsagn eller lover» (Tranøy, 2021). I tillegg er det ifølge Saunders et al. (2016, s. 144) en tredje tilnærming til teoriutvikling som er vel så vanlig innen forskning, kalt abduktiv tilnærming. Denne kjennetegnes ved et overraskende faktum/funn som utgjør selve konklusjonen, og søkes forklart gjennom å ta utgangspunkt i aktuelle premisser som kan forklare funnet (Saunders et al., 2016, s. 144, egen oversettelse).

Saunders et al. (2016, s. 145) tydeliggjør hvordan de ovenstående definisjonen skal forstås i praksis. Forfatterne angir at dersom man tar utgangspunkt i en teori og tester denne gjennom

---

antagelser formulert gjennom hypoteser, har man en deduktiv tilnærming. Induktiv tilnærming innebærer å starte med datainnsamling, og deretter utvikle teori med utgangspunkt i observasjonene (Saunders et al., 2016, s. 145). Forskningen har ikke bakgrunn i overraskende funn, og abduksjon anses følgelig ikke videre relevant.

Det foreligger studier både på hvordan kunstig intelligens kan benyttes i beslutningsprosessen og på hvordan KI-baserte beslutningsstøtteverktøy påvirker utøvelsen av profesjonelt skjønn. Et utvalg av disse er omtalt i kapittel 2, og disse er også utgangspunktet for utforming av hypotesene i kapittel 3. Følgelig har vi en deduktiv tilnærming.

### 4.1.2 Metodiske valg

Metodikken som lagt til grunn for forskningen vår er et resultat av flere valg foretatt underveis i prosessen. Det første valget er ifølge Saunders et al. (2016, s. 164) om det skal benyttes en kvantitativ-, kvalitativ metode eller en kombinasjon av disse. Kvantitativ/(kvalitativ) metode brukes ved innsamling og analyse av kvantitative<sup>10</sup>/(kvalitative<sup>11</sup>) data (Grønmo, 2020; 2021). Vi har i vår forskning benyttet oss av en kvantitativ metode ettersom vi fant det hensiktsmessig å tallfeste resultatene, for enklere å kunne stadfeste eventuelle forskjeller mellom gruppene<sup>12</sup>. Statistiske analyser av tallmessige størrelser gir i tillegg muligheten til å konkludere på hvorvidt det foreligger statistisk signifikante resultater, som gjør det mulig å forkaste nullhypotesen(e).

Valg av metode må også foretas mellom utforskende, beskrivende, forklarende, evaluerende eller en kombinasjon av disse (Saunders et al., 2016, s. 164). Vi begrenser drøftelsen her til argumentasjon rundt det foretrukne valget. Med bakgrunn i forskningsspørsmålet, og det faktum at vi ønsket å se på den kausale sammenhengen mellom variabler, gjør at vi havner innenfor et «forklarende» formål med vår forskning (Saunders et al., 2016, s. 176). Forfatterne angir videre at et forklarende formål innebærer å studere en situasjon, for å kunne forklare sammenhengen mellom variabler. Det var nettopp dette vi ønsket, hvilket er redegjort for under punkt 3.1.

---

<sup>10</sup> Data som foreligger i form av tall eller andre mengdetemer (Grønmo, 2021).

<sup>11</sup> Data som vanligvis foreligger i form av tekst (Grønmo, 2020).

<sup>12</sup> Redegjort for under punkt 4.1.4.

### 4.1.3 Forskningsstrategi

Forskningsstrategi «er den metodiske koblingen mellom egen filosofi og påfølgende valg av metoder for å samle inn og analysere data» (Denzin og Lincoln, 2011, i Saunders et al., 2016, s. 177, egen oversettelse). Det finnes en rekke ulike forskningsstrategier, hvor valget i stor grad avhenger av øvrig metodikk lagt til grunn. Saunders et al. (2016, s. 168) argumenterer imidlertid for at de primære forskningsstrategiene i forbindelse med kvantitativ metode er eksperiment og spørreundersøkelse. Formålet med et eksperiment er å finne sannsynligheten for at en endring i en uavhengig variabel påvirker den avhengige variabelen (Saunders et al., 2016, s. 178, egen oversettelse). Følgelig er forskningsstrategien vår et eksperiment. Dette er også ifølge Trotman (2001, s. 182) den vanligste strategien i forbindelse med å utforske revisor sine vurderinger.

Det finnes ulike typer eksperimentelle design, hvor et av disse er et klassisk eksperiment (Saunders et al., 2016, s. 179). Her blir det aktuelle utvalget<sup>13</sup> tilfeldig fordelt mellom en eller flere «eksperimentgruppe(r)» og en kontrollgruppe. I førstnevnte gruppe vil respondentene bli utsatt for en planlagt manipulasjon, hvilket ikke vil være tilfellet i sistnevnte gruppe. I vårt eksperiment har vi én kontrollgruppe og to eksperimentgrupper, visualisert gjennom et faktorielt design.

### 4.1.4 Faktorielt design

Fokuset ved faktorielle design er ifølge Trochim et al. (2016, s. 237) nærmest utelukkende på oppsettet av programmet eller manipulasjonen, komponentene og hoveddimensjonene. En definisjon av faktorielle design er «design som fokuserer på et program eller en manipulasjon, komponentene og hoveddimensjonene, og gjør det mulig å vurdere hvorvidt programmet har en effekt, hvorvidt ulike delkomponenter er effektive, og hvorvidt det finnes interaksjoner mellom effektene forårsaket av delkomponentene» (Trochim et al., 2016, s. 237, egen oversettelse).

I forkant av utarbeidelsen av det faktorielle designet, må man bestemme seg for hvor mange uavhengige variabler som skal inkluderes, og antall nivåer av de uavhengige variablene

---

<sup>13</sup> Se punkt 4.2.1

(Trotman, 2001, s. 184). Trochim et al. (2016, s. 238) poengterer viktigheten av begrepsbruk i forbindelse med beskrivelsen av elementene ved et faktorielt design. Uavhengige variabler skal omtales som faktorer, hvor nivåer er en underdimensjon av faktoren (Trochim et al., 2016, s. 238). I vårt design har vi to faktorer og to nivåer av disse:

		<b>Bruk/ikke bruk av KI-verktøy</b>	
<b>Forklaring av verktøyet</b>	(1) Ikke bruk/ikke forklaring	(2) Bruk/ikke forklaring	
	(3) IA	(4) Bruk/forklaring	

*Figur 2: 2x2 faktorielt design*

Dette resulterer altså i fire grupper og et 2x2 faktorielt design. Gruppe 1 utgjør kontrollgruppen hvor respondentene ikke utsettes for noe form for manipulasjon. I gruppe 2 inkluderes KI-verktøyet, men kun konklusjonen presenteres for respondenten. En ytterligere forklaring av KI-verktøyet gis til gruppe 4. Gruppe 3 har vi satt som inaktiv ettersom denne gruppen ikke vil være relevant å teste. Dette som følge av at en forklaring av verktøyet ikke vil være aktuelt når verktøyet ikke tas i bruk. I tilfeller som dette, hvor en eller flere celle(r) med hensikt ekskluderes, omtales som et «ukomplett faktorielt design» (Trochim et al., 2016, s. 247).

Hypotese 1 testes i utgangspunktet ved å sammenligne gruppe 1 og 2, ettersom manipulasjonen her vil være implementering av KI. Dersom vi ikke får signifikante resultater mellom disse to gruppene, vil vi bytte ut gruppe 2 med gruppe 4. Dette ettersom effekten av manipulasjonen i gruppe 2, forsterkes ytterligere i gruppe 4. Med utgangspunkt i drøftet teori, er vår antagelse at gruppe 2 og 4 vil redusere vurderingen av usikkerheten, sett mot gruppe 1. Hypotese 2 testes ved å sammenligne svarene relatert til skjønnsutøvelsen, mellom gruppe 2 og 4. I dette tilfellet forventer vi at respondentene vurderer usikkerheten som lavere i gruppe 4, ettersom tilliten til verktøyet her forventes å være høyere. Vi sammenligner også gruppe 2 og 4 i hypotese 3, men basert på svarene angående tillit til, og vektlegging av, kunstig intelligens i vurderingen. På denne måten får vi testet om resultatene fra hypotese 2 kan forstås ved endringer i algoritme-aversjon.

### 4.1.5 Between-/within subjects

Manipuleringen av de respektive uavhengige variablene kan gjøres enten «within subjects», «between subjects» eller ved en kombinasjon av disse (Trotman, 2001, s. 185). Ved between subjects vil hver respondent tilfeldig allokteres til ett nivå, mens man ved et within subjects vil eksponere alle respondentene til alle faktornivåene. En kombinasjon av disse, ved et 2x2 faktorielt design, innebærer å tildele ulike respondenter ett nivå innen den ene faktoren, men begge nivåene av den andre faktoren (Trotman, 2001, s. 185).

Valget mellom de ulike måtene å manipulere faktorene på, avhenger av en avveining mellom fordeler og ulemper ved de ulike metodene, samt omstendighetene rundt forskningsprosjektet. Charness et al. (2011, s. 2) peker på tre fordeler ved within subjects sett i forhold til between subjects: (1) intern validitet avhenger ikke av randomisering, (2) den statistiske kraften vil ofte være bedre og (3) metoden passer ofte bedre inn i et teoretisk perspektiv. I tillegg krever within subjects færre respondenter enn between subjects (Trotman, 2001, s. 185). En ulempe ved within subjects er imidlertid «demand effect», definert som at «respondentene i eksperimentet tolker (/gjennomskuer) eksperimentets intensjoner og endrer atferden tilsvarende, enten bevisst eller ikke» (Rosenthal, 1976, White, 1977, i Charness et al., 2011, s. 2, egen oversettelse). I tillegg vil undersøkelsen ved within subjects være mer omfattende, hvilket kan føre til frafall av respondenter underveis i besvarelsen (Trotman, 2001, s. 185).

Vårt valg av variabelmanipulasjon baserte seg blant annet på antagelser og erfaringer rundt utvalgsstørrelsen<sup>14</sup>. Et begrenset utvalg taler for et within-subjects design, av hensyn til å oppnå statistisk signifikante observasjoner. I forbindelse med utarbeidelsen av det faktorielle designet var vi inntil flere ulike gruppeantall, som ville gjort et within subjects design svært aktuelt. Det endelige faktorielle designet (2x2), samt at det ene nivået ikke var relevant å teste, gjorde imidlertid at et between subjects design kunne benyttes med tanke på den statistiske kraften. Dette så vi på som en fordel, ettersom vi ønsket å hensynta «demand» effekten i størst mulig grad. Grunnen var at vi hadde en presumsjon om at effekten vi ønsket å teste i eksperimentet, ville bli for åpenlys i et within subjects design.

---

<sup>14</sup> Se punkt 4.2.1 for en videre drøftelse av utvalget

---

## 4.2 Datainnsamling

### 4.2.1 Utvalg

Utvalg er «prosessen med å velge enheter fra en aktuell populasjon, for å kunne generalisere resultatene til populasjonen enhetene ble valgt fra» (Trochim et al., 2016, s. 80, egen oversettelse). Populasjonen utgjør mengden utvalget trekkes ut ifra (Trochim et al., 2016, s. 82). Ut ifra forskningsspørsmålet vårt vil alle praktiserende revisorer i Norge, som er med i revisjonsteamet eller på andre måter delaktig i vurderingen av fortsatt drift, utgjøre populasjonen vår. Metodisk teoretisk skulle vi da trukket et representativt utvalg av disse, for at resultatene for utvalget skal bli tilnærmet lik som for hele populasjonen (Trochim et al., 2016, s. 83). Dette ble vurdert å være noe omstendelig å få til, med tanke på tilgjengelig tid, begrensede ressurser og «busy season» i revisjonsbransjen på tidspunktet undersøkelsen ble distribuert.

På grunn av at praktiserende revisorer er en begrenset og kostbar ressurs, skal de ikke brukes med mindre det er nødvendig (Peecher & Solomon, 2001, s. 201, egen oversettelse). Det argumenteres i den forbindelse for at studenter kan fungere som surrogater for praktiserende revisorer (Peecher & Solomon, 2001, s. 199-201; Mortensen et al., 2012, s. 262). Følgelig utgjør den aktuelle populasjonen alle MRR-studenter, ettersom besvarelse av undersøkelsen krever innsikt i fagfeltet. Innenfor rammene av omfanget til masteroppgaven, fant vi et ikke-sannsynlighetsutvalg hensiktsmessig. Utvalgstypen kjennetegnes ved at utvalget ikke trekkes tilfeldig (Trochim et al., 2016, s. 86).

Ikke-sannsynlighetsvalg kjennetegnes ved en iboende fare for at man ender opp med et systematisk skjevt utvalg, hvilket medfører at man ikke kan generalisere resultatene fra utvalg til populasjon (Jacobsen, 2015, s. 302). Dette er naturligvis uheldig i forbindelse med gjennomføringen av et eksperiment, hvor det sentrale imidlertid vil være å anerkjenne den begrensede generaliserbarheten. Utvalget ble nærmere bestemt valgt gjennom et bekvemmelighetsutvalg, hvilket innebærer å trekke ut «de som det er lettest å få tak i» (Jacobsen, 2015, s. 302). Respondentene som ble valgt ut er derfor i hovedsak MRR-studenter ved Norges Handelshøyskole, men også nyutdannede revisormedarbeidere i vårt nettverk.



## 4.2.2 Utforming av eksperimentell undersøkelse

Utgangspunktet for undersøkelsen vår var et case om et flyselskap. Caset er inspirert fra tidligere eksamens- og gruppeoppgaver i faget MRR451 Revisjon II ved NHH. Dette ettersom faget problematiserer revisor sine vurderinger knyttet til fortsatt drift-forutsetningen. Formålet med caset var å presentere respondentene for et saksforhold hvor det forelå usikkerhet knyttet til enhetens evne til fortsatt drift. Det var ikke meningen med en forhåndsbestemt konklusjon, men at respondentene skulle svare ut ifra egne vurderinger. Caset kan leses i sin helhet under «vedlegg 1». Valget om å beskrive et flyselskap er basert på bransjens natur. Betydelige konkurransekrefter, stadige konjunktursvingninger og kapitalintensiteten mener vi utgjør et godt grunnlag for et fortsatt drift-case. Vi valgte for øvrig å eksplisitt angi at selskapet drev innen interkontinental godsfrakt for å unngå eventuelle assosiasjoner til den nasjonale flybransjen, på grunn av eventuell påvirkning dette kunne medføre.

Alle respondentene fikk informasjonen som nettopp spesifisert, hvor gruppe 2 og 4 i ulik grad fikk ytterligere informasjon om et KI-beslutningsstøtteverktøy. Dette utgjorde manipulasjonen i eksperimentet. Vi ønsket å kontrollere for effekten av uoppmerksomhet eller ignoranse ved å inkludere en manipulasjonssjekk, og stilte respondentene derfor følgende spørsmål:

*Alle gruppene: Fikk du informasjon om at et KI-verktøy ble tatt i bruk?*

*Gruppe 2 og 4: Fikk du informasjon om prediksjonsevnen til KI-verktøyet?*

Etter at respondentene var introdusert for caset, ble det umiddelbart gitt tre spørsmål relatert til den avhengige variabelen. Spørsmålene ble gitt på samme side som caset for å ha tilgang til informasjonen ved behov. Respondentene ble gitt rangordnede spørsmål, kjennetegnet ved å alltid forholde seg til ytterpunkter på en skala (Jacobsen, 2015, s. 262). I forbindelse med utformingen av svaralternativene måtte vi ta stilling til hvor mange svarnivå som skulle gis. Dette både med tanke på antallet isolert sett, men også om det skulle gis en middelvei. Forskning antyder at mellom fem og ni alternativer er det optimale antallet (Preston & Colman, 2000, s. 11). Ytterpunktene til svaralternativene ble formulert «lite sannsynlig» og «svært sannsynlig», eller tilsvarende. Dette kalles ofte for en Likert-skala (Malt & Grønmo, 2020). Dernest måtte vi altså ta stilling til om skalaen skulle inneholde en middelvei, eller eventuelt en «forced-choice response scale» (Trochim et al., 2016, s. 154). Vi valgte sistnevnte for å øke

---

sannsynligheten for å måle en effekt blant observasjonene. Vi fant følgelig seks alternativer å være formålstjenlig.

Spørsmål relatert til alder, kjønn, erfaring og masterretning var også noe respondentene måtte svare på i undersøkelsen. Alle respondentene ble spurt om dette. Gruppe 2 og 4 ble gitt tilleggsspørsmål utover dette, for eksplisitt å kunne måle effekten av algoritme-aversjon. Vi stilte spørsmål som gjorde det mulig å måle i hvilken grad algoritmen ble vektlagt i vurderingen, samt tilliten til denne. Dette gjorde det også mulig å kontrollere for ulik algoritme-aversjon, samt måle effekten av eventuelle endringer i algoritme-aversjon mellom gruppe 2 og 4.

### **4.2.3 Pilottest**

Enhver undersøkelse bør pilottestes før den distribueres for å få et inntrykk av tidsbruken på undersøkelsen, samt om respondentene forstår instruksjonene og spørsmålene (Thabane et al., 2010, s. 2). Ved å gjennomføre en pilottest reduserer man risikoen for målingsfeil og dermed også sannsynligheten for å påbegynne en tidkrevende og kostbar datainnhenting, som ikke er tilpasset formålet (Trochim et al., 2016, s. 118). Vi gjennomførte pilottesten på fem sisteårs MRR-studenter, som også har noe arbeidserfaring innen revisjon. Hensikten med dette var i tillegg å kontrollere manipulasjonssjekkene og om påstandene besvarte hypotesene.

Pilottesten ble gjennomført i to omganger. I første omgang gjennomførte vi pilottesten på en annen mastergruppe med samme veileder, ettersom de har kjennskap til anvendt metode. Etter å ha observert svarene deres og mottatt tilbakemeldinger, valgte vi å moderere informasjonen angående KI-verktøyet i gruppe 4. Nedkorting ble gjort for å få en tydeligere effekt av å benytte kunstig intelligens i beslutningstakingen.

I den neste omgangen testet vi på tre nye medstudenter, der den ene hadde tre års erfaring innen revisjon. Vi observerte svarene og om effekten av endringene fra den første pilottesten, stemte i henhold til våre antagelser. Vi fikk tilbakemelding om at caset virket realistisk, og at spørsmålene var enkle å forstå. Alle bestod manipulasjonssjekkene, og syntes de virket hensiktsmessige.

## 4.2.4 Praktisk gjennomføring

Datagrunnlaget for eksperimentet ble hentet inn ved hjelp av en spørreundersøkelse utformet og distribuert i Qualtrics, som tilbyr nødvendige funksjoner. Vi benyttet blant annet randomiseringsfunksjonen, som er avgjørende for å måle årsak-virkningsforholdet vi ønsket å avdekke (Trochim et al., 2016, s. 221). I praksis betyr dette at ved tre gjennomførte undersøkelser, vil respondentene havne i hver sin gruppe.

Vi slo av funksjonen som samler inn IP-adresser. Dette for å gjøre undersøkelsen anonym, som innebærer at «det skal være umulig å koble informasjon med opplysninger om enkeltpersoners identitet» (Jacobsen, 2015, s. 50). Anonymisering kan bidra til å sikre at respondentene svarer med utgangspunkt i egne tanker og vurderinger heller enn det antatt riktige svaret (APS, 2019a; Trotman, 2001, s. 185). Vi aktiverte også en funksjon som gjorde at respondentene måtte svare på spørsmålene som ble stilt, for å unngå ufullstendige svar. For at manipulasjonssjekkene skulle fungere, ble disse spørsmålene plassert på siden etter caset, uten mulighet til å trykke seg tilbake.

Undersøkelsen ble delt med aktuelle respondenter gjennom digitale distribusjonskanaler, samt ved formidling i forelesninger for MRR-studenter. En utfordring med innsamling av data er stadig publikasjon av undersøkelser, som reduserer villigheten til å delta. Dette ble hensyntatt ved å kontakte aktuelle respondenter direkte, fremfor mer generiske forespørsler. Personlig rettede forespørsler samt deltakelse i forelesninger, viste seg å være helt nødvendig.

## 4.3 Analysemetode

### 4.3.1 Dataanalyse

For statistiske analyser ønsker man å finne sannsynligheten for om resultatene er realistiske eller kun tilfeldige. Dette gjør man ved å sammenligne sannsynligheten med det valgte signifikansnivået for forskningen. For samfunnsvitenskapelige eksperimenter er det vanlig å benytte en kritisk verdi lik 5%. Dersom resultatene fra eksperimentet har en p-verdi som overstiger signifikansnivået, vil ikke resultatene være statistisk signifikante, og vi kan ikke forkaste nullhypotesen. Videre betyr det at det er 5% sannsynlighet for type I feil, altså at en feilaktig forkaster nullhypotesen (Trochim et al., 2016, s. 283).

---

Vi benyttet analyseprogrammet IBM SPSS for analysen. Da vi avsluttet undersøkelsen i Qualtrics ble dataen eksportert direkte til SPSS. Dermed reduserte vi risikoen for menneskelige feil i overføringen, som sikrer påliteligheten til dataene. For å vurdere hvorvidt vi kunne forkaste nullhypotesene, gjennomførte vi som utgangspunkt en t-test ettersom vi ikke sammenligner mer enn to populasjoner for hver hypotese. Analysemetoden gjør det mulig å teste om det er statistisk signifikante forskjeller i gjennomsnittet til gruppene (Trochim et al., 2016, s. 317; Laerd Statistics, u.å.a).

Vi operasjonaliserte den avhengige variabelen gjennom tre spørsmål der det er nødvendig at det utøves profesjonelt skjønn. Vi vurderte signifikansen av spørsmålene enkeltvis, og samlet som en indeks. I tilfeller der man benytter en indeks, er det avgjørende at spørsmålene er konsistente og måler det samme. Vi benytter Cronbach's Alpha for å vurdere reliabiliteten til indeksene, som anses pålitelig når verdien er over 0,7 (Saunders et al., 2016, s. 451). Vi benytter én indeks for profesjonelt skjønn og én for algoritme-aversjon i forbindelse med henholdsvis hypotese 1 og 2, og hypotese 3.

### **4.3.2 Forutsetninger for analysen**

For å benytte en t-test, må den underliggende dataen tilfredsstillende seks forutsetninger (Laerd Statistics, u.å.a). (1) Den avhengige variabelen måles kontinuerlig, (2) de uavhengige variablene består av to eller flere kategoriske variabler, (3) observasjonene er uavhengige fra hverandre, (4) det er ingen signifikante utligger, (5) den avhengige variabelen er tilnærmet normalfordelt og (6) det må være homogenitet.

De tre første forutsetningene oppfylles gjennom foretatte metodiske valg som redegjort for. Forutsetning 4 kontrollerer vi gjennom boksplottet av observerte verdier, som angir om det foreligger observasjoner som avviker signifikant fra øvrige observasjoner. Vi tester normalfordeling gjennom en Shapiro-Wilk test. Ved brudd på denne forutsetningen, gjennomfører vi istedenfor en Mann-Whitney U test. Denne testen har fire forutsetninger, hvorav de tre første er tilsvarende som for en t-test, og den fjerde at dataene ikke er normalfordelte (Laerd Statistics, u.å.b). Testen ser dermed om det er signifikante forskjeller mellom gruppene når den avhengige variabelen ikke er normalfordelt. I tilfeller der forutsetning 5 holder, hensyntas forutsetning 6 gjennom Leven's test. Ved brudd på en eller flere av forutsetningene, drøftes implikasjonene av dette i forbindelse med den eller de aktuelle hypotesen(e) i kapittel 5.

## 4.4 Forskningsetikk og personvern

Det var frivillig å delta i eksperimentet, og deltakerne kunne når som helst trekke seg under gjennomføringen. Vi sørget for at alle retningslinjer for personvern ble oppfylt. Ved innsamling av personvernsopplysninger er forskningen meldepliktig til Norsk senter for forskningsdata (NSD). Vi deaktiverte alle funksjoner i Qualtrics som samler inn personopplysninger, og stilte ingen spørsmål som gjør det mulig å identifisere respondenten. Deltakernes anonymitet ble dermed sikret, og forskningen var ikke meldepliktig til NSD.

## 4.5 Reliabilitet og validitet

Reliabilitet og validitet er to svært sentrale forhold i forbindelse med å vurdere kvaliteten til et eksperiment. Reliabilitet handler om påliteligheten og konsistensen til målingene, mens validiteten viser til i hvilken grad undersøkelsen måler det den skal måle (Trochim et al., 2016, s. 115; Saunders et al., 2016, s. 202). Kriteriene henger tett sammen, da reliabiliteten til en undersøkelse er sentralt for å vurdere validiteten. Om undersøkelsen ikke regnes som reliabel, vil det altså ikke være noe poeng å vurdere validiteten (Trochim et al., 2016, s. 127; Grenness, 2012, s. 106).

Med reliabiliteten evaluerer man om undersøkelsen gir konsistente målinger på ulike tidspunkt og under forskjellige omstendigheter (Saunders et al., 2016, s. 451). Tilfeldige og systematiske feil, også kalt målingsfeil, er kilder som kan svekke reliabiliteten til undersøkelsen. Tilfeldige feil er faktorer som tilfeldig påvirker målingene, eksempelvis dagsformen til respondenten. Systematiske feil er faktorer som systematisk påvirker målingene, eksempelvis feil relatert til metoden. (Trochim et al., 2016, s. 117).

Vi har tatt følgende forholdsregler for å redusere risikoen for målingsfeil. Respondentene blir tilfeldig fordelt i de tre casegruppene for å ta høyde for tilfeldige feil, og vi oppnår da en normalfordeling der tilfeldige feil går mot null (Trochim et al., 2016, s. 17). På samme måte vil også de underliggende forskjellene i respondentene fordeles jevnt mellom alle gruppene, slik at vi kan ta utgangspunkt i «the law of chance», der vi ignorerer andre mulige faktorer enn den manipulerede (Peecher & Solomon, 2001, s. 196). Som følge av begrenset tid og ressurser, har vårt beskjedne utvalg imidlertid en effekt på om vi har oppnådd «the law of chance». For å redusere sannsynligheten for systematiske feil gjennomførte vi en pilottest på flere

---

medstudenter, som redegjort for i kapittel 4.2.3. Som nevnt i kapittel 4.2.2 ønsket vi et case der respondenten selv skulle konkludere ut fra den gitte informasjonen. Det medfører en viss risiko for at respondentene vurderer informasjonen ulikt, som kan svekke reliabiliteten til undersøkelsen. Likevel har vi gjennom pilottestene fått en tydelig indikasjon på at respondentene ikke oppfatter informasjonen ulikt. Vi benytter også flere spørsmål for å måle profesjonelt skjønn og algoritme-aversjon, som reduserer risikoen for systematiske feil, samt gir en mer nøyaktig måling av effekten (Trochim et al., 2016, s. 118; Saunders et al., 2016, s. 451). Vi kontrollerer som nevnt for den interne konsistensen til spørsmålene som en indeks, ved å benytte Cronbach's Alpha (Saunders et al., 2016, s. 451).

Reliabilitet er en sentral og nødvendig, men ikke tilstrekkelig, forutsetning for kvaliteten og validiteten til en undersøkelse (Saunders et al., 2016, s. 451; Grenness, 2012, s. 106). Validitet handler om gyldighet og relevans, og om man klarer å måle det man faktisk har til hensikt å måle (Jacobsen, 2015, s. 352; Saunders et al., 2016, s. 202). Validitet er et omfattende begrep med flere delkategorier (Grenness, 2012, s. 112). Vi vil videre fokusere på intern og ekstern validitet.

Den første validitetsformen vi vil redegjøre for er intern validitet. Eksperimentet vil ha intern validitet dersom endringen i den avhengige variabelen kun skyldes manipulasjonene i den uavhengige variabelen (Trotman, 2001, s. 186; Grenness, 2012, s. 117). For å styrke den interne validiteten ble respondentene tilfeldig fordelt mellom de tre gruppene i eksperimentet. Dette er et effektivt tiltak ettersom det reduserer risikoen for at det oppstår systematiske forskjeller mellom gruppene (Peecher & Solomon, 2001, s. 185-186). Når undersøkelsen gjennomføres digitalt øker sannsynligheten for frafall, som kan føre til at gruppene blir systematisk forskjellige, og dermed ikke sammenlignbare (Trotman, 2001, s. 187). I hovedsak distribuerte vi eksperimentet digitalt, men også fysisk i forelesning. Fysisk tilstedeværelse styrker den interne validiteten, ved at vi kunne kontrollere at respondentene gjennomførte eksperimentet individuelt, tidsbruken til respondentene og at de fullførte eksperimentet (Trotman, 2001, s. 189). Vi kunne ikke kontrollere undersøkelsen for de som gjennomførte den digitalt, men vi oppfordret alle til å gjennomføre undersøkelsen selvstendig.

Vi inkluderte en kontrollvariabel for å øke sannsynligheten for at årsakssammenhengen mellom den avhengige variabelen og de uavhengige variablene, skyldes endringer i sistnevnte (Trotman, 2001, s. 186). Manipulasjonssjekkene sikret også at deltakerne fikk med seg endringene i de uavhengige variablene, som styrker den interne validiteten.

Videre vil vi redegjøre for ekstern validitet, som vil si at resultatene også vil være gjeldende i sammenhenger utenfor eksperimentet (Trochim et al., 2016, s. 83; Saunders et al., 2016, s. 151). Det handler altså om generaliserbarheten av funnene til den gjeldende populasjonen. Tiltaket som gir best forutsetning for å oppnå ekstern validitet, er tilfeldig utvalg (Trochim et al., 2016, s. 83). Ettersom vi har benyttet et bekvemmelighetsutvalg, som gjort rede for under punkt 4.2.1, svekkes den eksterne validiteten ved at utvalget kan være systematisk forskjellig fra populasjonen (Jacobsen, 2015, s. 365; Trochim et al., 2016, s. 86). Vi har som kjent benyttet MRR-studenter som surrogater, og det kan diskuteres om de er representative for praktiserende revisorer. Tidligere forskning viser imidlertid at dette har en begrenset påvirkning på den eksterne validiteten (Peecher & Solomon, 2001, s. 199-201; Mortensen et al., 2012, s. 262).

Det oppstår ofte en avveining mellom intern og ekstern validitet. Høy grad av intern validitet gjør ofte eksperimentet særegent, som reduserer mulighetene for generalisering (Grenness, 2012, s. 118). Peecher og Solomon (2001, s. 198) advarer mot å øke ekstern validitet på bekostning av intern validitet, ettersom en undersøkelse uten intern validitet ikke kan generaliseres. Som følge av at et eksperimentelt forskningsdesign gir resultater som ikke er egnet for generalisering (Trochim et al., 2016, s. 252; Saunders et al., 2016, s. 138), har vi fokusert på den interne validiteten.

---

## 5. Resultater

Resultatene av dataanalysen presenteres i kapittel 5. Rådata ble eksportert direkte fra Qualtrics til SPSS, og vi måtte derfor begynne med å «rydde» disse. Dernest var det mulig å analysere svarene fra undersøkelsen, som angis med bakgrunn i hver enkelt hypotese. Resultatene diskuteres i det påfølgende kapittelet.

### 5.1 Rensing av data

Som beskrevet i kapittel 4.2.1, bestod utvalget vårt av revisjonsstudenter ved NHH, i tillegg til yrkesaktive i vårt nettverk. Totalt var det 71 personer som åpnet undersøkelsen, men til sammen 19 svar ble slettet som følge av at svarene ikke var fullstendige. Totalt mottok vi altså 52 fullstendige svar, som gir en frafallsrate på 27%. Ettersom eksperimentet i hovedsak ble gjennomført i ikke-kontrollerte omgivelser, og distribuert digital, er det ikke uvanlig med en høy frafallsrate (Dandurand et al., 2008, s. 428). Frafallet gjorde at vi fikk en systematisk forskjell i antall respondenter i gruppene (tabell 1), noe som kan svekke den interne validiteten. Vi antar likevel at gruppene vil være sammenlignbare.

Som nevnt i kapittel 4.2.1, besto utvalget vårt både av revisjonsstudenter og nyutdannede revisormedarbeidere. Ettersom vi er kjent med at ikke alle i sistnevnte gruppe har gått MRR, stilte vi spørsmål om arbeidserfaring og masterretning. Det var ikke ønskelig at erfaring skulle føre til forskjeller mellom gruppene, og følgelig satt vi krav om maks tre års relevant arbeidserfaring. Dersom det var respondenter som ikke hadde gått MRR, stilte vi i tillegg krav om minst to års arbeidserfaring for å sikre tilstrekkelig faglig kompetanse. Kravene var ment å underbygge sammenlignbarheten mellom gruppene og følgelig styrke den interne validiteten. Tre respondenter oppfylte ikke kompetansekravet og ble derfor fjernet.

For å øke sannsynligheten for at respondentene satt seg inn i caset og vurderte informasjonen kritisk, studerte vi gjennomføringstiden. En av respondentene brukte 1,5 minutter på hele undersøkelsen, og vi vurderte at denne personen ikke hadde gjort en tilstrekkelig innsats. Respondenten ble følgelig fjernet fra analysen. I tillegg var svarene til en annen respondent vesentlig forskjellig fra de øvrige, og manglet også intuisjon. Følgelig fjernet vi også denne respondenten, og vi hadde dermed 47 fullstendige svar før vi vurderte manipulasjonssjekken.



### 5.1.1 Manipulasjonssjekk

Manipulasjonssjekken viste seg å være nyttig, ettersom noen respondenter ikke fikk med seg all tilleggsinformasjonen. Dette gjaldt tre respondenter i både gruppe 2 og 4, hvis svar ble fjernet i sin helhet som følger. Svarene på de to spørsmålene som utgjorde manipulasjonssjekken, tyder på at caset ikke har blitt lest i det hele tatt, manglende observasjonsevne og/eller tid. Dette basert på avgitte svar, hvor et av de ikke-beståtte svarene mente å ha fått informasjon om prediksjonsevnen til KI-verktøyet, men likevel ingen informasjon om KI.

En svakhet ved manipulasjonssjekken er imidlertid at respondenter som ikke har lest gjennom caset, eventuelt lest det for fort, har tippet riktig på manipulasjonssjekken. Dette anses ikke å være usannsynlig, gitt 50% sannsynlighet for å svare «riktig» på spørsmålet. Svakheten er som nevnt forsøkt kontrollert for ved å vurdere tidsbruken til respondentene, noe som antyder hvorvidt caset er lest eller ikke.

Mangelen på ikke-beståtte svar i gruppe 1, gjør at denne skiller seg noe fra de to andre gruppene. En naturlig forklaring er at det bare ble gitt ett spørsmål som manipulasjonssjekk til denne gruppen. Mer informasjon ble i tillegg gitt gruppe 2 og 4, noe som naturligvis kompliserer manipulasjonssjekken. Vi mener imidlertid at effekten av randomiseringen fortsatt er i behold. Ubalansen i antall respondenter i hver gruppe kan imidlertid svekke validiteten, og vi vil derfor være oppmerksomme på eventuelle implikasjoner i videre analyse.

## 5.2 Deskriptiv statistikk

Deskriptiv statistikk er beskrivende statistikk som brukes for å gi informasjon om det grunnleggende i datasettet (Grennes, 2012, s. 175; Trochim et al., 2016, s. 291). Statistikken er hjelpelig ved å gjøre store mengder data forståelig (Trochim et al., 2016, s. 291). Den er også nyttig for å undersøke om det foreligger systematiske forskjeller mellom gruppene, som kan medføre at endringer i den avhengige variabelen skyldes noe annet enn manipulasjonen i den uavhengige variabelen. En av de vanligste metodene for beskrivende statistikk er frekvenstabeller (Trochim et al., 2016, s. 291).

		<b>Gruppe 1</b>	<b>i %</b>	<b>Gruppe 2</b>	<b>i %</b>	<b>Gruppe 4</b>	<b>i %</b>	<b>Total</b>
<b>Antall</b>		17		11		13		41
<b>Kjønn</b>	Mann	9	53 %	7	64 %	8	62 %	59 %
	Kvinne	8	47 %	4	36 %	5	38 %	41 %
<b>Alder</b>	20-25	7	41 %	8	73 %	3	23 %	44 %
	26-30	8	47 %	3	27 %	9	69 %	49 %
	Over 30	2	12 %	0	0 %	1	8 %	7 %
<b>Masterretning</b>	MRR	14	82 %	11	100 %	12	92 %	90 %
	MØA	3	18 %	0	0 %	1	8 %	10 %
<b>Arbeidserfaring</b>	Nei	4	24 %	4	36 %	5	38 %	32 %
	<1 år	6	35 %	5	45 %	5	38 %	39 %
	1-2 år	1	6 %	1	9 %	1	8 %	7 %
	2-3 år	6	35 %	1	9 %	2	15 %	22 %

Tabell 1: Deskriptiv statistikk

Tabellen viser frekvensen av kontrollvariablene totalt og innenfor de individuelle gruppene, både som numeriske verdier og prosentvis. Respondentene ble spurt om hvilket kjønn de var. Svaralternativene var «mann», «kvinne» og «annet», og ingen av respondentene svarte «annet». Kjønnfordelingen i de tre gruppene er tilnærmet likt som for utvalget. Dette er også tilfellet for masterretning, hvor majoriteten har utdanning innen revisjon.

De fleste av respondentene befinner seg innenfor aldersintervallet 20-30 år. Dette er naturlig ettersom vi i hovedsak benytter studenter i utvalget. Det relativt unge utvalget har også en innvirkning for den begrensede erfaringen til respondentene. Ettersom erfaringsnivået er noe ulikt på tvers av gruppene, vil skjevfordelingen kunne skape problemer for den interne validiteten. Utover dette fordeler de demografiske variablene seg jevnt over de tre gruppene.

### 5.3 Hypotesetesting

Før vi analyserer resultatene for hver hypotese, kontrollerer vi at forutsetningene som angitt i kapittel 4.3.2 er oppfylt. Vi begynner med en boksploTTanalyse av resultatene for å kartlegge eventuelle utliggere. Dersom det skulle foreligge slike, men resultatene ikke påvirkes, vil vi beholde disse. Dette grunnet det i utgangspunktet begrensede utvalget. Ved brudd på normalfordelingsforutsetningen<sup>15</sup>, utfører vi en Mann-Whitney U test. Dersom denne ikke endrer resultatene, vil vi se vekk fra bruddet på normalfordeling og presentere tall fra t-testen

<sup>15</sup> Shapiro-Wilk test, se vedlegg 3

om ikke annet spesifikt angis. Forutsetningen om homogenitet testes ved Leven's test, hvis brudd gjør at vi presenterer p-verdi ved heterogenitet dersom denne endrer resultatene.

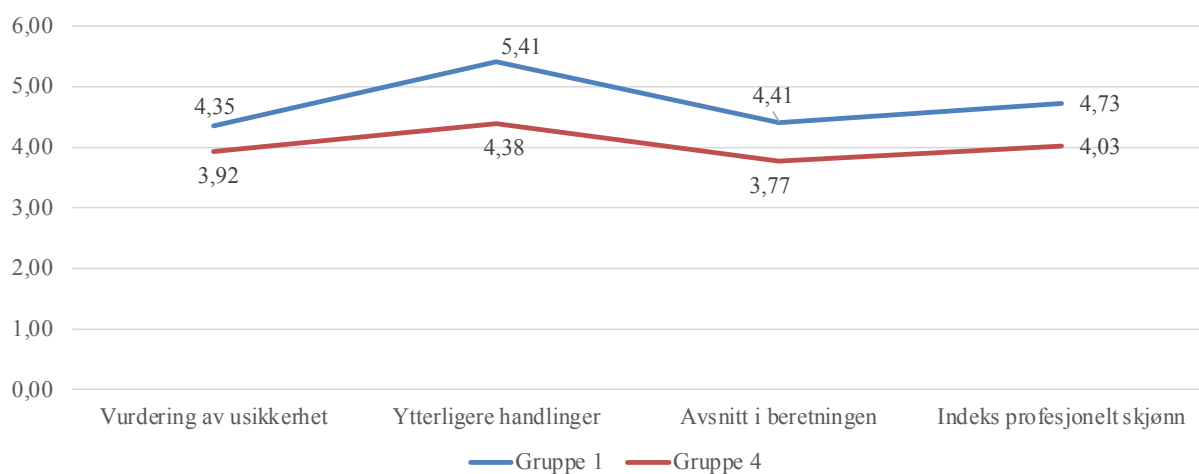
Indeksen for profesjonelt skjønn og algoritme-aversjon har en Cronbach's Alpha (vedlegg 2) på henholdsvis 0,833 og 0,743. Disse anses derfor som pålitelige, og vil benyttes i videre analyse.

### 5.3.1 Forutsetninger analyse

Delkapittelet er avgrenset til tilfeller der det foreligger brudd på forutsetningene, for indeksene. Boksplottet (vedlegg 5) for profesjonelt skjønn-indeksen til gruppe 2 indikerer at denne inneholder en utligger. Ettersom fjerning av dette datapunktet ikke endrer resultatene, beholdes det. Profesjonelt skjønn-indeksen til gruppe 2 bryter også forutsetningene om normalfordeling og homogenitet. Verken Mann-Whitney U- eller Leven's testen endrer resultatene (vedlegg 5).

### 5.3.2 Resultater hypotese 1: Implementering av KI i vurderingen

Hypotese 1 tester om tilgangen på et KI-verktøy som konkluderer med at fortsatt drift-forutsetningen er til stede, fører til at revisor vurderer at det foreligger mindre usikkerhet til enhetens evne til fortsatt drift. Utgangspunktet vårt var en sammenligning av gruppe 1 og 2. Forskjellen i gjennomsnittsverdier mellom disse to var imidlertid neglisjerbar, og vi fant det derfor mer aktuelt å se nærmere på gruppe 1 og 4.



Figur 3: Gjennomsnittssvar for gruppe 1 og 4 – profesjonelt skjønn

Gjennomsnittssvarene tyder på at revisor, gjennom en skjønnsmessig vurdering, vurderer at det foreligger mindre usikkerhet til enhetens evne til fortsatt drift, ved tilgang på et KI-basert beslutningsstøtteverktøy. Dette gjelder både for spørsmålene enkeltvis og som en indeks, og er i tråd med angitt teori. Vi gjennomfører en t-test for å kartlegge om observasjonen er signifikant. Deskriptiv statistikk og p-verdier fra testen er angitt i tabell 2.

	Deskriptiv statistikk		Independent T-test Sig. (1-tailed)
	Gruppe	Antall	
Vurdering av usikkerhet	1 4	17 13	0,165
Ytterligere handlinger	1 4	17 13	0,015
Avsnitt i beretningen	1 4	17 13	0,123
Indeks profesjonelt skjønn	1 4	17 13	0,056

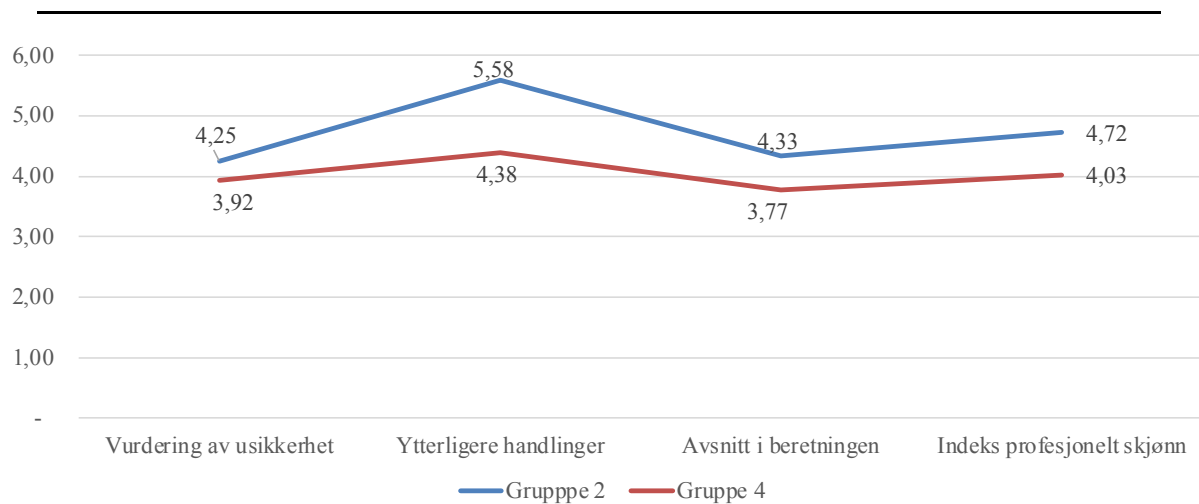
Tabell 2: Deskriptiv statistikk og p-verdier – hypotese 1

Resultatene viser at indeksen har en p-verdi ( $p = .056$ ) som er høyere enn signifikansnivået på 0,05. Vi kan dermed ikke konkludere på at det er noen signifikant forskjell i skjønnsutøvelsen. Av svarene sett enkeltvis ser vi at det kun er «ytterligere handlinger» som er signifikant ( $p = .015$ ).

**H<sub>0</sub> beholdes.**

### 5.3.3 Resultater hypotese 2: Effekten av forklaring på vurderingen av saksforholdet

I hypotese 2 tester vi om en forklaring av verktøyet (gruppe 4) fører til at revisor vurderer at det foreligger mindre usikkerhet til enhetens evne til fortsatt drift, sammenlignet med tilfellet hvor respondenten kun blir presentert for konklusjonen (gruppe 2).



Figur 4: Gjennomsnittssvar for gruppe 2 og 4 – profesjonelt skjønn

Gjennomsnittssvarene tyder på at revisor vurderer at det foreligger mindre usikkerhet til enhetens evne til fortsatt drift, når det gis en forklaring av verktøyet, for spørsmålene enkeltvis og samlet. Dette er i tråd med angitt teori. Vi gjennomfører en t-test for å kartlegge om observasjonen er signifikant. Deskriptiv statistikk og p-verdier fra testen er angitt i tabell 3.

	Deskriptiv statistikk		Independent T-test
	Gruppe	Antall	Sig. (1-tailed)
Vurdering av usikkerhet	2 4	11 13	0,225
Ytterligere handlinger	2 4	11 13	0,016
Avsnitt i beretningen	2 4	11 13	0,142
Indeks profesjonelt skjønn	2 4	11 13	0,061

Tabell 3: Deskriptiv statistikk og p-verdier – hypotese 2

Resultatene viser at indeksen har en p-verdi ( $p = .061$ ) som er høyere enn signifikansnivået på 0,05. Vi kan dermed ikke konkludere på at det er noen signifikant forskjell i skjønnsutøvelsen. Av svarene sett enkeltvis ser vi at det også her kun er «ytterligere handlinger» som er signifikant ( $p = .016$ ).

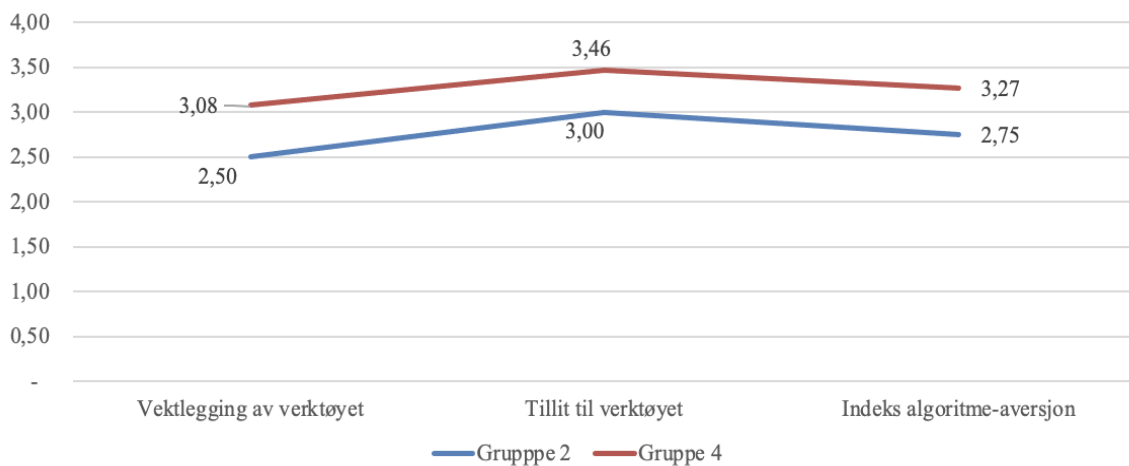
For å undersøke hvorvidt kontrollvariabelen algoritme-aversjon påvirker resultatene fra t-testen, kunne en ANCOVA-analyse vært hensiktsmessig. Denne innehar imidlertid samme forutsetninger som en t-test, hvilket det foreligger brudd på som diskutert. Vi finner det følgelig mer aktuelt å gjennomføre en ikke-parametrisk test også i dette tilfellet, nærmere

bestemt en Quade's ANCOVA (Vedlegg 5). For indeksen profesjonelt skjønn finner vi at analysen ikke endrer resultatene fra t-testen ( $p = .312$ ). Derimot finner vi at «ytterligere handlinger» ikke lenger er signifikant ( $p = .179$ ) når vi kontrollerer for algoritme-aversjon.

$H_0$  beholdes.

### 5.3.4 Resultater hypotese 3: Algoritme-aversjon

I den tredje hypotesen tester vi om en forklaring av KI-verktøyet fører til redusert algoritme-aversjon. Vi sammenligner svarene til gruppe 2 og 4, ettersom forskjellen mellom disse utelukkende er forklaring av verktøyet.



Figur 5: Gjennomsnittssvar for gruppe 2 og 4 – algoritme-aversjon

Gjennomsnittssvarene tyder på at revisor sin algoritme-aversjon reduseres når det gis en forklaring av verktøyet, for spørsmålene enkeltvis og samlet. Dette er i tråd med angitt teori. Vi gjennomfører en t-test for å kartlegge om observasjonen er signifikant. Deskriptiv statistikk og p-verdier fra testen er angitt i tabell 4.

	Deskriptiv statistikk		Independent T-test Sig. (1-tailed)
	Gruppe	Antall	
Vektlegging av verktøyet	2 4	11 13	0,120
Tillit til verktøyet	2 4	11 13	0,048
Indeks algoritme-aversjon	2 4	11 13	0,053

Tabell 4: Deskriptiv statistikk og p-verdier – hypotese 3

Resultatene viser at indeksen for algoritme-aversjon ikke er signifikant ( $p = .053$ ). Vi kan dermed ikke konkludere på at det er en signifikant forskjell i algoritme-aversjonen. Av svarene sett enkeltvis ser vi at «tillit til verktøyet» er signifikant ( $p = .048$ ).

**H<sub>0</sub> beholdes.**

---

## 6. Diskusjon og konklusjon

Under diskuteres resultatene som presentert i forrige kapittel, før vi konkluderer på forskningsspørsmålet.

### 6.1 Diskusjon av hypoteser

#### 6.1.1 Hypotese 1

$H_0$  beholdes: *Revisor vurderer at det foreligger uendret eller mer usikkerhet til enhetens evne til fortsatt drift, ved tilgang på et KI-verktøy som konkluderer med at fortsatt drift-forutsetningen er til stede.*

Funnene våre gir ikke grunnlag for å konkludere i tråd med studien til Zhang et al. (2020) som kom frem til at KI-baserte beslutningsstøtteverktøy øker kvaliteten på beslutningene. Forutsatt at respondentene har et snev av tillit til algoritmen, skulle man følgelig forvente at usikkerheten ble vurdert lavere. Dette kan naturligvis skyldes måten undersøkelsen er utformet på, da denne er en simplifisert fremstilling av vurderingen. Som kjent er vurderingen av forutsetningen om fortsatt drift svært sammensatt og kompleks, hvor en konstruert sammenligning i de fleste tilfeller ikke vil være representativt for et faktisk tilfelle. Følgelig kan det tenkes at en mer virkelighetsnær fremstilling, som fører til økt situasjonell forståelse for respondentene, ville ført til andre resultater.

Som nevnt var gjennomsnittssvarene mellom gruppe 1 og 2 nærmest identiske. Dette tilsier i praksis at en konklusjon fra et beslutningsstøtteverktøy basert på KI alene, ikke endrer vurderingene av fortsatt drift-forutsetningen. En tilbakeholdende bruk av revisjonsbevis som revisor ikke har kunnskap om, er i tråd med revisjonsstandardene som tilsier at revisor i ustrakt grad skal sikre påliteligheten av bevisene. Tolkningen som her lagt til grunn tyder altså på at implementering av KI i sin enkleste form, i den aktuelle beslutningsprosessen, ikke vil ha verdi.

Utfallet av hypotese 1 kan tyde på at den iboende usikkerheten ved en vurdering av fortsatt drift generelt sett, er for betydelig til at man ser en signifikant effekt av KI-verktøyet ( $p = .056$ ). Dette kan dels forklares ved at en teoretisk sett objektiv oppgave, også preges av en viss grad av subjektivitet (Castelo et al., 2019, s. 817). Resultatene tilsier imidlertid at det er en



signifikant reduksjon ( $p = .015$ ) i sannsynligheten for å utføre ytterligere handlinger, fra gruppe 1 til 4. Isolert sett fører det til en mer kostnadseffektiv revisjon, og kan forstås som at KI-verktøyet erstatter en del av de tiltenke handlingene.

Sett i ettetid var spørsmålet om å legge til et avsnitt i beretningen noe mindre treffende. ISA 570 angir som nevnt at et avsnitt skal legges til kun i de tilfeller der forutsetningen om fortsatt drift kan legges til grunn, det foreligger en vesentlig usikkerhet og det er gitt tilstrekkelig opplysninger. Det er svært sjeldent at forutsetningen om fortsatt drift ikke kan legges til grunn for regnskapet, noe det kan forventes at respondentene er kjent med. Tatt i betraktning at usikkerheten ikke er signifikant forskjellig mellom gruppene ( $p = .165$ ), og opplysninger i regnskapet er holdt utenfor undersøkelsen i sin helhet, finner vi det naturlig at det ikke er redusert signifikant sannsynlighet ( $p = .123$ ) for å legge til et avsnitt. Det er imidlertid uklart om respondentene har observert denne detaljen, eller om spørsmålet likevel kan betraktes som en proxy.

Resultatene indikerer imidlertid en tendens til endring i utøvelsen av profesjonelt skjønn, og kan sies å være marginalt signifikante (APS, 2019b). Dersom resultatene kan betraktes som signifikante, ville det antageligvis kun ført til bedret kvalitet og en mer kostnadseffektiv revisjon, i den grad finanstilsynet hadde stilt seg positiv til bruk av kunstig intelligens i denne type vurderinger. Det må antas at eventuelle kostnadsbesparelser som måtte følge av bruk av KI, ved dagens aktuelle regulering, på sikt ville ha blitt redusert av finanstilsynets bemerkninger. Resultatene fra gruppe 1 og 2, sammenlignet med 1 og 4, tyder på at revisor utøver hensiktsmessig profesjonelt skjønn.

### **6.1.2 Hypotese 2**

*H<sub>0</sub> beholdes: Revisor vurderer at det foreligger uendret eller mer usikkerhet til enhetens evne til fortsatt drift, når det gis en forklaring av KI-verktøyet sammenlignet med når det ikke gis en forklaring.*

Resultatene samlet sett tyder på at en forklaring av verktøyet ikke vil ha signifikant påvirkning ( $p = .061$ ) på vurderingen av fortsatt drift. Tatt i betraktning at den aktuelle vurderingen bærer preg av betydelig usikkerhet, vil algoritme-aversjonen øke (Dietvorst & Bharti, 2020, s. 1313). Det kan derfor tenkes at kunnskap om algoritmen isolert sett, ikke var tilstrekkelig til å motvirke dette (Burton et al., 2020, s. 223). En ytterligere forklaring kan være at vurderingen av fortsatt drift innehar for høy grad av subjektivitet til at redegjørelsen av algoritmen vil ha

---

noen effekt. Dette strider til en viss grad mot Castelo et al. (2019) sine funn, om at kunnskap om treffsikkerheten til algoritmen, reduserer algoritme-aversjon også i subjektive vurderinger.

Vi finner dog også her marginalt signifikante resultater fra t-testen, hvilket tyder på at revisor tenderer til å benytte det KI-baserte beslutningsstøtteverktøyet når det gis en forklaring av det. Dette er i tråd med funnene i hypotese 1, om at en forklaring av verktøyet er en nødvendighet for å ta det i bruk. Til tross for at forklaringen ikke var tilstrekkelig til å kunne forkaste nullhypotesen, er det altså mye som tyder på at informasjonen som ble gitt likevel ikke er irrelevant. Dersom revisor til en viss grad hensyntar verktøyet i fortsatt drift-vurderingene, vil det kunne øke kvaliteten av disse (Chi & Shen, 2022, s. 16). Økt kvalitet bidrar til å opprettholde tilliten til revisor i kapitalmarkedene, ved at både type I og II feil knyttet til fortsatt drift-forutsetningen forhåpentligvis reduseres.

Av resultatene fra t-testen individuelt sett, er det som nevnt kun ytterligere handlinger som er statistisk signifikant forskjellig mellom gruppene. Resultatene indikerer dermed at revisor anser verktøyet som pålitelig, og vil benytte det som et revisjonsbevis. Grunnen kan tenkes å være hensynet til notoriteten ved revisjonen, hvor informasjon som underbygger påliteligheten til verktøyet vil kunne dokumenteres. Øvrige drøftelser rundt denne observasjonen er gitt i delkapittel 6.1.1. Resultatene er imidlertid ikke signifikante når vi kontrollerer for algoritme-aversjon, med p-verdi for indeks og ytterligere handlinger på henholdsvis 0,312 og 0,179. Dette skyldes sannsynligvis at effekten av forklaring uteblir som følge av at kontrollvariabelen også påvirkes av forklaringen. Forskjeller i algoritme-aversjon mellom gruppene ses nærmere på i hypotese 3.

### 6.1.3 Hypotese 3

$H_0$  beholdes: *Revisor sin algoritme-aversjon er uendret eller høyere når det gis en forklaring av KI-verktøyet.*

Basert på indeksen for algoritme-aversjon gir funnene fra analysen ikke grunnlag ( $p = .053$ ) til å konkludere på at aversjonen reduseres når det gis en forklaring av beslutningsstøtteverktøyet. Tidligere teori stadfester at algoritme-aversjon er et svært komplekst teorem, som kan være vanskelig å redusere. Dette vil være spesielt gjeldende i situasjoner som bærer preg av høy usikkerhet, hvor det følgelig kan tenkes at forklaring alene ikke er tilstrekkelig til å gi en signifikant reduksjon i aversjonen. En annen drøftelse av at vi ikke så noen effekt på algoritme-aversjonen, er at respondentene oppfattet caset som for enkelt.

Dette kan ses i sammenheng med en av forklaringene på algoritme-aversjon, om at mennesker har en overdreven tro på egne ferdigheter. Dersom det er tilfellet, vil respondentene naturligvis ikke hensynta konklusjonen til beslutningsstøtteverktøyet i særlig grad.

Svarene tyder imidlertid på at en forklaring har en signifikant påvirkning på tilliten til algoritmen. Dette er i tråd med tidligere diskutert teori, om at informasjon om treffsikkerhet til algoritmen og dens iboende begrensninger, øker tilliten (Yin et al., 2019, s. 6; Zhang et al., 2020, s. 295). Den økte tilliten forstås som at respondentene er villige til å ta i bruk KI-baserte beslutningsstøtteverktøy, og at det ikke er tilliten til algoritmen som vil forsinke implementeringen. Følgelig kan det tyde på at det er holdningene til tilsynsmyndigheter som forsinke bruken av slike verktøy (Eilifsen, 2020, s. 76).

Vi observerte ingen signifikant forskjell mellom gruppene, hva gjelder vektlegging av algoritmen. Dette er i tråd med hypotese 2. Funnet tyder på at en forklaring av verktøyet ikke påvirker i hvilken grad det vektlegges i vurderingen, hvilket er naturlig sett i sammenheng med finanstilsynet sin uttrykte skepsis. Det antas for øvrig at respondentene er kjent med den lite utstrakte bruken av algoritmer i dagens revisjonspraksis. En forklaring kan da være at respondentene anså det som en risiko å være de første som tok i bruk et slikt verktøy. Dette bygger på et kjent psykologisk teorem om at man følger majoriteten selv i tilfeller dette viser seg å være feil (Crano & Seyranian, 2007, s. 573). En tidligere studie viser at nettopp informasjon om andre, pålitelige individers vellykkede bruk av algoritmen, kan være et effektivt tiltak for å redusere aversjonen (Alexander et al., 2018, s. 287).

Finanstilsynet sin uttrykte skepsis til bruken av KI kan blant annet skyldes at algoritmene vurderer saksforholdet på ett gitt tidspunkt. Revisjonsstandardene krever oppmerksomhet fra revisor gjennom hele revisjonen, og følgelig foreligger det en inkonsistens mellom revisor og algoritmen på dette området. At problematikken rundt innsikten i algoritmen sin vurdering leder til skepsis, sammenfaller med en av fallgruvene ved skjønnsutøvelse om at mangel på begrunnelsesplikt fører til vilkårlighet. Vilkaarlige beslutninger på området, er naturligvis lite ønsket. Det er derfor ikke urimelig at finanstilsynet først vil evaluere bruken av kunstig intelligens.

---

## 6.2 Implikasjoner for praksis

Generelt sett tyder resultatene våre på at en implementering av et KI-basert beslutningsstøtteverktøy i fortsatt drift-vurderingen, ikke vil ha vesentlig effekt. Vi har ikke grunnlag for å hevde at bruken av kunstig intelligens i den aktuelle prosessen, vil føre til verken en mer kostnadseffektiv revisjon eller øke kvaliteten av den. Slik omstendighetene foreligger for øyeblikket, herunder aktuell regulering på området og regulatoriske myndigheter tatt i betraktning, kan det se ut til at investeringene på området ikke vil gi umiddelbar avkastning for revisjonsselskapene. Vi finner imidlertid at revisor vurderer det mer aktuelt å ta i bruk KI-verktøy, dersom vedkommende gis mulighet til å tilegne seg kunnskap om det. Dette under forutsetning av at verktøyet blant annet har høy treffsikkerhet og er utviklet av en troverdig aktør.

At tilliten til algoritmen øker ved forklaring, men at den likevel ikke vektlegges i egen vurdering, impliserer at standardsettere og tilsynsmyndigheter begrenser digitaliseringen av revisjonsbransjen. Dette tilsier imidlertid at den strenge reguleringen av bransjen fungerer etter sin hensikt, all den tid påliteligheten til kunstig intelligens ikke kan etterprøves.

## 6.3 Konklusjon

Formålet med denne utredningen var å undersøke om et KI-basert beslutningsstøtteverktøy påvirker utøvelsen av profesjonelt skjønn i en fortsatt drift-vurdering, og om kunnskap om algoritmen reduserer algoritme-aversjon.

Resultatene fra eksperimentet tilsier at et KI-basert beslutningsstøtteverktøy ikke har påvirkning på vurderingen av fortsatt drift ved 5% signifikansnivå. Vi fant imidlertid at en forklaring av KI-verktøyet har en signifikant effekt for uttrykt tillit til algoritmen. Sett i sammenheng er det mulig å redusere revisor sin algoritme-aversjon uttrykt ved tillit, men at bransjen er sitt ansvar om etterlevelse av lov og god revisjonsskikk bevisst.

## 6.4 Begrensninger

Omstendighetene ved utredningen har ført med seg noen begrensninger ved forskningen. En av disse er knyttet til utvalget. Tidsbegrensningen for forskningen, samt at aktuelle respondenter ikke er fullt så villige til å delta, gjør at antall respondenter er noe begrenset. Et begrenset utvalg og den uønskede skjevfordelingen av respondenter mellom gruppene, gjorde det nødvendig å forutsette sammenlignbarhet. Dette gjør imidlertid studiet sårbar for å trekke feilslutninger, som følge av at p-verdiene er sensitive for svar fra enkeltrespondenter. Hensynet til å i det hele tatt få tak i respondenter gjør i tillegg at undersøkelsen må utformes så kort og presis som mulig, for å unngå et betydelig frafall. Dette kan ses på som et optimaliseringsproblem, hvor man må veie hensynet til situasjonsforståelse mot omfanget av undersøkelsen.

Hensynet til innhold i undersøkelsen, gjorde at vi måtte være selektive til hva som skulle inkluderes av informasjon. Spesielt gjaldt dette forklaringen av KI-verktøyet. Basert på drøftet teori kan det tenkes at mer informasjon om KI-verktøyet, herunder informasjon om selve beslutningsprosessen, kunne påvirket resultatene. Det er nemlig ikke gitt at kunnskap om algoritmen alene vil øke tilliten i betydelig grad, men samspillet mellom flere forhold. Eksempelvis kunnskap om begrensninger ved algoritmen, egen interaksjon og andres vellykkede bruk av tilsvarende verktøy.

Mangelfull tilgang på tid og ressurser førte som nevnt til at vi valgte et bekvemmelighetsutvalg. Utvalgstypen resulterte i at majoriteten av respondentene er NHH-studenter, hvilket medfører redusert ekstern validitet. Optimalt sett burde vi antageligvis hatt et forholdsmessig utvalg fra alle skolene i Norge som tilbyr MRR.

Det er i oppgaven antatt at studenter kan benyttes som surrogater for praktiserende revisorer, med utgangspunkt i tidligere angitt teori. Studenter har likevel ikke praktisk erfaring i særlig grad, hvilket er ett av kriteriene for profesjonelt skjønn per definisjon. Følgelig kan det tenkes at studenter i noe mindre grad er egnet som surrogater for praktiserende revisorer i dette tilfellet.

---

## 6.5 Forslag til videre forskning

Omfanget av forskningsområdet vi har valgt å se nærmere på, har gjort avgrensninger underveis nødvendig. I den forbindelse peker vi her på et utvalg aktuelle temaer for fremtidig forskning, som vi har bemerket oss underveis. Vi finner det blant annet interessant å se nærmere på bruk av kunstig intelligens i forbindelse med vurderinger/beslutninger som krever bruk av profesjonelt skjønn, men med mindre konsekvenser og usikkerhet enn i en fortsatt drift-vurdering.

Et annet aktuelt eksperiment vil være å teste om rekkefølgen til vurderingene har betydning. Altså om vurderingene påvirkes av om konklusjonen til KI-verktøyet gis før eller etter egen konklusjon. Dette er imidlertid definert som «recency bias», hvor det allerede foreligger en del tidligere forskning. «Recency bias»-effekten kan tenkes å ha større effekt i praksis, ettersom bevisinnsamlingen for fortsatt drift gjør seg gjeldende gjennom hele revisjonen.

Aktuelt kan det også være å teste i hvor stor grad finanstilsynet sitt syn på bruken av kunstig intelligens, påvirker vurderingene og vektleggingen av verktøyet. Skyldes en tilbakeholdende bruk av verktøyet i en fortsatt drift-vurdering utelukkende finanstilsynet, eller er det andre faktorer som er vel så viktige? I forlengelse av et slikt eksperiment kunne man naturligvis kartlagt hvilke eventuelle faktorer som veier tyngst, eller påvirker vurderingen i det hele tatt.

Et siste forslag til videre forskning er å teste en motsatt hypoteseretning. Altså tilfellet der det foreligger mindre usikkerhet i den aktuelle situasjonen, og algoritmen konkluderer med at fortsatt-drift forutsetningen ikke kan legges til grunn.

## Litteraturliste

- Abdolmohammadi, M. J. (1999). A comprehensive taxonomy of audit task structure, professional rank and decision aids for behavioral research. *Behavioral Research in Accounting*, 11, 51-92. <https://www.proquest.com/scholarly-journals/comprehensive-taxonomy-audit-task-structure/docview/203299937/se-2?accountid=37265>
- Agnew, H. (2016). Auditing: Pitch battle. *Financial Times*.  
<https://www.ft.com/content/268637f6-15c8-11e6-9d98-00386a18e39d>
- AICPA. (2017). *Audit Data Analytics (ADAs) Can transform Audits; New AICPA Guide Will Help Auditors Apply ADA Techniques*.  
<https://us.aicpa.org/press/pressreleases/2017/audit-data-analytics-new-aicpa-guide-will-help-auditors-apply-ada-techniques>
- Alexander, V., Blinder, C. & Zak, P. J. (2018). Why trust an algorithm? Performance, cognition, and neurophysiology. *Computers in Human Behavior*, 89, 279–288.  
<https://doi.org/10.1016/j.chb.2018.07.026>
- APS. (2019a). *Under time pressure, people tell us what we want to hear*.  
<https://www.psychologicalscience.org/news/releases/under-time-pressure-people-tell-us-what-we-want-to-hear.html>
- APS. (2019b). *How marginal are “marginally significant” p-values?*  
<https://www.psychologicalscience.org/observer/how-marginal-are-marginally-significant-p-values>
- Ashton, R. H. & Ashton, A. H. (1995). *Judgment and decision-making research in accounting and auditing*. Cambridge University Press.
- Banko, M. & Brill, E. (2001). Scaling to very very large corpora for natural language disambiguation. *Association for Computational Linguistics*, 26-33.  
<https://doi.org/10.3115/1073012.1073017>
- Bernhoft, A. C., Kvifte, S. S. & Røsok, K. O. (2018). *IFRS i Norge: en håndbok*. (8. utgave). Fagbokforlaget.
- Birnbaum, M. H. (1976). Intuitive Numerical Prediction. *The American Journal of Psychology*, 89(3), 417-429. <https://doi.org/10.2307/1421615>
- Birnbaum, M. H. & Stegner, S. E. (1979). Source credibility in social judgment: Bias, expertise, and the judge's point of view. *Journal of Personality and Social Psychology*, 37(1), 48–74. <https://doi.org/10.1037/0022-3514.37.1.48>

- 
- Braut, G. S. (2021, 24. juni). *type I-feil*. SNL. [https://snl.no/type\\_I-feil](https://snl.no/type_I-feil)
- Brennan, B., Flynn, M. & Baccala, M. (2017). *Artificial Intelligence Comes to Financial Statement Audits*. CFO. <https://www.cfo.com/accounting-tax/2017/02/artificial-intelligence-audits/>
- Brynjolfsson, E. & McAfee, A. (2019). *Artificial Intelligence. The Business of Artificial Intelligence* (1. utg). Harvard Business Review Press
- Burgess, A. (2018). *The Executive Guide to Artificial Intelligence: How to identify and implement applications for AI in your organization* (1. utg). Palgrave MacMillan
- Burton, J. W., Stein, M. & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, 33(10), 220-239. <https://doi.org/10.1002/bdm.2155>
- Carson, E., Fargher, N. L., Geiger, M. A., Lennox, C. S., Raghunandan, K. & Willekens, M. (2013). Audit Reporting for Going-Concern Uncertainty: A Research Synthesis. *Auditing: A Journal of Practice & Theory*, 32(1). 353-384. <https://doi.org/10.2308/ajpt-50324>
- Castelo, N., Bos, M. W. & Lehmann, D. R. (2019). Task-Dependent Algorithm Aversion. *Journal of Marketing Research*, 56(5). 809-825. <https://doi.org/10.1177/0022243719851788>
- Chao, C. Y., Chang, T. C., Wu, H. C., Lin, Y. S., & Chen, P. C. (2016). The interrelationship between intelligent agents' characteristics and users' intention in a search engine by making beliefs and perceived risks mediators. *Computers in Human Behavior*, 64, 117–125. <https://doi.org/10.1016/j.chb.2016.06.031>
- Charness, G., Gneezy, U. & Kuhn, M. A. (2011). Experimental methods: Between subject and within subject design. *Journal of Economic Behavior & Organization*, 81(1), 1-8. <https://doi.org/10.1016/j.jebo.2011.08.009>
- Chi, D. J. & Chu, C. C. (2021). Artificial Intelligence in Corporate Sustainability: Using LSTM and GRU for Going Concern Prediction. *Sustainability 2021*, 13(21), 1-18. <https://doi.org/10.3390/su132111631>
- Chi, D. J. & Shen, Z. (2022). Using Hybrid Artificial Intelligence and Machine Learning Technologies for Sustainability in Going-Concern Prediction. *Sustainability 2022*, 14(3), 1-18. <https://doi.org/10.3390/su14031810>
- Colson, E. (2019). *What AI-Driven Decision Making Looks Like*. Harvard Business Review. <https://hbr.org/2019/07/what-ai-driven-decision-making-looks-like>



- Commerford, B. P., Dennis, S. A., Joe, J. R. & Ulla, J. W. (2021). Man Versus Machine: Complex Estimates and Auditor Reliance on Artificial Intelligence. *Journal of Accounting Research*, 60(1), 171-201. <https://doi.org/10.1111/1475-679X.12407>
- Crano, W. D. & Seyranian, V. (2007). Majority and Minority Influence. *Social and Personality Psychology Compass*, 1(1), 572-589. <https://doi.org/10.1111/j.1751-9004.2007.00028.x>
- Dandurand, F., Shultz, T. R. & Onishi, K. H. (2008). Comparing online and lab methods in a problem-solving experiment. *Behavior Research Methods* 40(2), 428-434. <https://doi.org/10.3758/BRM.40.2.428>
- Datatilsynet. (2018). *Kunstig intelligens og personvern*. <https://www.datatilsynet.no/globalassets/global/dokumenter-pdf-skjema-ol/rettigheter-og-plikter/rapporter/rapport-om-ki-og-personvern.pdf>
- Deloitte. (2017). *Delivering smarter audits: Insights through innovation*. <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/audit/us-audit-smarter-audits-dynamic-insights-through-innovation.pdf>
- Deloitte. (2019). *Managing the black box of artificial intelligence (AI)*. <https://www2.deloitte.com/us/en/pages/advisory/articles/black-box-artificial-intelligence.html>
- Deloitte. (2021). *Teknologien snur tradisjonelt revisjonsarbeid på hodet*. <https://www2.deloitte.com/no/no/pages/audit/articles/Analytics-revisjon.html>
- Dietvorst, B. J. & Bharti, S. (2020). People Reject Algorithms in Uncertain Decision Domains Because They Have Diminishing Sensitivity to Forecasting Error. *Psychological Science*, 31(10), 1302-1314. <https://doi.org/10.1177/0956797620948841>
- Dietvorst, B. J., Simmons, J. P. & Massey, C. (2015). Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err. *Journal of Experimental Psychology: General*, 144(1), 114-126. <https://doi.org/10.1037/xge0000033>
- Dietvorst, B. J., Simmons, J. P. & Massey, C. (2018). Overcoming Algorithm Aversion: People Will Use Imperfect. *Management Science*, 64(3), 1155-1170. <https://doi.org/10.1287/mnsc.2016.2643>
- Eilifsen, A., Kinserdal, F., Messier Jr, W. F. & McKee, T. E. (2020). An Exploratory Study into the Use of Audit Data Analytics on Audit Engagements. *Accounting Horizons*, 34(4), 75-103. <https://doi.org/10.2308/HORIZONS-19-121>

- 
- Einhorn, H. J. (1986). Accepting Error to Make Less Error. *Journal of Personality Assessment*, 50(3), 387-395. [https://doi.org/10.1207/s15327752jpa5003\\_8](https://doi.org/10.1207/s15327752jpa5003_8)
- Elkins, A. C., Dunbar, N. D., Adame, B. & Nunamaker, J. F. (2013). Are Users Threatened by Credibility Assessment Systems? *Journal of Management Information Systems*, 29(4), 249-262. <https://doi.org/10.2753/MIS0742-1222290409>
- ESMA. (2020). *Implications of the COVID-19 outbreak on the half-yearly financial reports*. [https://www.esma.europa.eu/sites/default/files/library/esma32-63-972\\_public\\_statement\\_on\\_half-yearly\\_financial\\_reports\\_in\\_relation\\_to\\_covid-19.pdf](https://www.esma.europa.eu/sites/default/files/library/esma32-63-972_public_statement_on_half-yearly_financial_reports_in_relation_to_covid-19.pdf)
- EY. (2021). *Revisjonstjenester*. [https://www.ey.com/no\\_no/audit/services](https://www.ey.com/no_no/audit/services)
- Finanstilsynet. (2020). *Betydning av COVID-19 utbryddet på halvårsrapporteringen for noterte foretak*. <https://www.finanstilsynet.no/nyhetsarkiv/nyheter/2020/betydning-av-covid-19-utbruddet-pa-halvarsrapporteringen-for-notere-foretak/>
- Fjørtoft, L. E. (2018). Digitalisering og disruptjon i revisjonsbransjen. *Revisjon og regnskap*, 1, 24–26. <https://www.revregn.no/journal/2018/1/revisjon-1-2018-922>
- Fossheim, H. J. & Ingierd, H. (2015). *Etisk skjønn i forskning*. Universitetsforlaget. <https://doi.org/10.18261/9788215025162-2015>
- Gorner, J. & Sweeney, A. (2020). For years Chicago police rated the risk of tens of thousands being caught up in violence. That controversial effort has quietly been ended. *Chicago Tribune*. <https://www.chicagotribune.com/news/criminal-justice/ct-chicago-police-strategic-subject-list-ended-20200125-spn4kjmrxrh4tmktdjckhtox4i-story.html>
- Grenness, T. (2012). *Hvordan kan du vite om noe er sant?: veiviser i forsknings- og utredningsarbeid for studenter* (2. utg). Cappelen Damm Akademisk.
- Grønmo, S. (2020). *Kvalitativ metode*. SNL. [https://snl.no/kvalitativ\\_metode](https://snl.no/kvalitativ_metode)
- Grønmo, S. (2021, 7. november). *Kvantitativ metode*. SNL. [https://snl.no/kvantitativ\\_metode](https://snl.no/kvantitativ_metode)
- Handeland, Ø., Schwencke, H. R., Stenheim, T. & Baksaas, K. M. (2020). *Regnskapsloven med kommentarer*. Gyldendal Akademisk.
- Hao, K. (2019). *AI is sending people to jail—and getting it wrong*. MIT Technology Review. <https://www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai/>
- Hardesty, L. (2017). *Explained: Neural networks*. MIT News Office. <https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414>
- Hengstler, M., Enkel, E. & Duelli, S. (2016). Applied artificial intelligence and trust—The case of autonomous vehicles and medical assistance devices. *Technological*

- 
- Forecasting and Social Change*, 105, 105-120.  
<https://doi.org/10.1016/j.techfore.2015.12.014>
- Highhouse, S. (2008). Stubborn Reliance on Intuition and Subjectivity in Employee Selection. *Industrial and Organizational Psychology*, 1(3), 333-342.  
<https://doi.org/10.1111/j.1754-9434.2008.00058.x>
- IAASB. (2009a). *ISA 200 Overordnede mål for den uavhengige revisor og gjennomføringen av en revisjon i samsvar med de internasjonale revisjonsstandardene*.  
<https://revisorforeningen.no/globalassets/fag/standarder-og-veiledninger/revisjonsstandardene/pr-05022021/isa-200-n-0121.pdf>
- IAASB. (2009b). *ISA 530 Stikkprøver i revisjon*.  
<https://www.revisorforeningen.no/globalassets/fag/standarder-og-veiledninger/revisjonsstandardene/isa-530-stikkprover-i-revisjon.pdf>
- IAASB. (2016a). *ISA 570 Fortsatt drift*.  
<https://www.revisorforeningen.no/globalassets/fag/standarder-og-veiledninger/revisjonsstandardene/isa-570-revidert-fortsatt-drift.pdf>
- IAASB. (2016b). *ISA 705 Modifikasjoner i konklusjonen i den uavhengige revisors beretning*. <https://www.revisorforeningen.no/globalassets/fag/standarder-og-veiledninger/revisjonsstandardene/isa-705-revidert-modifikasjoner-i-konklusjonen-i-den-uavhenige-revisors-beretning.pdf>
- IAASB. (2019). *Proposed Strategy for 2020- 2023 and Work Plan for 2020-2021*.  
<https://www.ifac.org/system/files/publications/files/IAASB-Strategy-for-2020-2023-V6.pdf>
- IASB. (2005). *IAS 1 Presentasjon av finansregnskap*. <https://app.dib.no/standard/ias-1-presentation-av-finansregnskap/%7B95728EF1-CCBD-4FF1-960D-9F632C2558A8%7D>
- IBM. (2020). *Neural Networks*. <https://www.ibm.com/cloud/learn/neural-networks>
- Israni, E. T. (2017). When an Algorithm Helps Send You to Prison. *The New York Times*.  
<https://www.nytimes.com/2017/10/26/opinion/algorithm-compas-sentencing-bias.html>
- Issa, H., Sun, T. & Vasarhelyi, M. A. (2016). Research Ideas for Artificial Intelligence in Auditing: The Formalization of Audit and Workforce Supplementation. *Journal of Emerging Technologies in Accounting*, 13(2), 1–20. <https://doi.org/10.2308/jeta-10511>

- 
- Jacob, N., Burton, C., Hale, R., Jones, A., Lloyd, A., Rafferty, A. M. & Allen, D. (2021). Pro-judge study: Nurses' professional judgement in nurse staffing systems. *Journal of Advanced Nursing*, 77, 4226-4233. <https://doi.org/10.1111/jan.14921>
- Jacobsen, D. I. (2015). *Hvordan gjennomføre undersøkelser?: Innføring i samfunnsvitenskapelig metode* (3. utgave). Cappelen Damm.
- Jan, C. (2021). Using Deep Learning Algorithms for CPAs' Going Concern Prediction. *Information*, 12(2), 1-22. <https://doi.org/10.3390/info12020073>
- Johnsen, A. & Kvaal, E. (1998). *Regnskapsloven: kommentarer til lov av 17. juli 1998 nr 56 om årsregnskap m.v.* Cappelen Akademisk Forlag.
- Jussupow, E., Benbasat, I. & Heinzl, A. (2020). Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion. *In Proceedings of the 28th European Conference on Information Systems (ECIS)*. 1-16. [https://aisel.aisnet.org/ecis2020\\_rp/168](https://aisel.aisnet.org/ecis2020_rp/168)
- Kahneman, D., Rosenfield, A. M., Gandhi, L. & Blaser, T. (2016). Noise: How to overcome the high, hidden cost of inconsistent decision making. *Harvard Business Review*, 94, 38-46. <https://hbr.org/2016/10/noise>
- Kahneman, D. & Klein, G. (2009). Conditions for Intuitive Expertise A Failure to Disagree. *American Psychologist*, 64(6), 515-526. <https://doi.org/10.1037/a0016755>
- Kaplan, A. & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1), 15-25. <https://doi.org/10.1016/j.bushor.2018.08.004>
- Kennedy, M., Ferrell, L. & LeClair, D. (2001). Consumers' trust of salesperson and manufacturer: an empirical study. *Journal of Business Research*, 51(1), 73-86. [https://doi.org/10.1016/S0148-2963\(99\)00039-9](https://doi.org/10.1016/S0148-2963(99)00039-9)
- Kinserdal, F. (2017). NHH skal forske på digitalisering i revisjonsbransjen. *Magma*, 76-86. <https://www.nhh.no/globalassets/centres/digaudit/publications/magma-2017-spesialnummer-nhh-skal-forske-pa-digitalisering-i-revisjon-finn-kinserdal.pdf>
- Kjellevoll, K. (2019). Revisors vurderinger ved usikkerhet om fortsatt drift. *Forskning og revisjon*, (2), 30-31. <https://www.revregn.no/asset/pdf/2019/2-30-1.pdf>
- Kokina, J. & Davenport, T. H. (2017). The Emergence of Artificial Intelligence: How Automation is Changing Auditing. *Journal of emerging technologies in accounting*, 14(1), 115-122. <https://doi.org/10.2308/jeta-51730>

- KPMG. (2018). *Teknologi øker revisjonskvaliteten og gir kunder merverdi*.  
<https://home.kpmg/no/nb/home/nyheter-og-innsikt/2018/11/teknologi-okere-revisjonskvaliteten-og-gir-kunden-mer-verdi.html>
- KPMG. (2019). *Artificial intelligence in Control with Watson OpenScale*.  
<https://www.kpmg.us/alliances/kpmg-ibm/ai-in-control-watson-openscale.html>
- Laerd Statistics. (u.å.a). Independent t-test using SPSS Statistics.  
<https://statistics.laerd.com/spss-tutorials/independent-t-test-using-spss-statistics.php>
- Laerd Statistics. (u.å.b). Mann-Whitney U Test using SPSS Statistics.  
<https://statistics.laerd.com/spss-tutorials/mann-whitney-u-test-using-spss-statistics.php>
- Lee, J. D. & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors*, 46(1), 50-80. [https://doi.org/10.1518/hfes.46.1.50\\_30392](https://doi.org/10.1518/hfes.46.1.50_30392)
- Lewis, J. D. & Weigert, A. (1985). Trust as a Social Reality. *Social Forces*, 63(4), 967–985.  
<https://doi.org/10.1093/sf/63.4.967>
- Libby, R. (1981). *Accounting and human information processing: Theory and applications*. Prentice-hall, Inc.
- Lien, L. E. (2012). Har regnskapsloven § 4-5 samme innhold som IAS 1? *Revisjon og regnskap*, (1), 23-25. [https://www.revregn.no/journal/2012/1/rr1-12b-698/Fortsatt\\_drift-forutsetningen](https://www.revregn.no/journal/2012/1/rr1-12b-698/Fortsatt_drift-forutsetningen)
- Lim, J. S. & O'Connor, M. (1995). Judgemental adjustment of initial forecasts: Its effectiveness and biases. *Journal of Behavioral Decision Making*, 8(3), 149–168.  
<https://doi.org/10.1002/bdm.3960080302>
- Longoni, C., Bonezzi, A. & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research*, 46(4). 629–650.  
<https://doi.org/10.1093/jcr/ucz013>
- Lorentzen, H. (2005). *Skjønn versus regler - om skjønnsutøvelsens vilkår og konsekvenser*. (NIBR-rapport 2005:7). Norsk institutt for by- og regionforskning.  
[https://kudos.dfo.no/files/eec/eec5c926146391b2479da88da93c3d46a8064755a732b71e06f4b3577b1db963/Skjonn\\_versus\\_regler.pdf](https://kudos.dfo.no/files/eec/eec5c926146391b2479da88da93c3d46a8064755a732b71e06f4b3577b1db963/Skjonn_versus_regler.pdf)
- Madhavan, P. & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: an integrative review. *Theoretical Issues in Ergonomics Science*, 8(4). 277-301. <https://doi.org/10.1080/14639220500337708>
- Malt, U. & Grønmo, S. (2020). *Likert-skala*. SNL. <https://snl.no/Likert-skala>

- 
- Mortensen, T., Fisher, R. & Wines, G. (2012). Students as surrogates for practicing accountants: Further evidence. *Accounting Forum*, 36(4), 251-265.  
<https://doi.org/10.1016/j.accfor.2012.06.003>
- Moxander, J. & Axente, M. (2021). Ethics-based auditing of automated decision-making systems: intervention points and policy implications. *AI & Society*, 1-19.  
<https://doi.org/10.1007/s00146-021-01286-x>
- Munoko, I., Brown-Liburd, H. & Vasarhelyi, M. (2020). The Ethical Implications of Using Artificial Intelligence in Auditing. *Journal of Business Ethics*, 167, 209-234.  
<https://doi.org/10.1007/s10551-019-04407-1>
- Nogler, G. E. (1995). The Resolution of Auditor Going Concern Opinions. *Auditing*, 14(2), 54–73. <https://www.proquest.com/scholarly-journals/resolution-auditor-going-concern-opinions/docview/216732451/se-2?accountid=37265>
- Norsk Regnskapsstiftelse. (2003). *NRS 13 Usikre forpliktelser og betingede eiendeler*.  
<https://www.regnskapsstiftelsen.no/wp-content/uploads/2008/06/NRS-13-Usikre-forpliktelse-og-betingede-eiendeler-2013.pdf>
- Norsk Regnskapsstiftelse. (2021, 01. oktober). *NRS ber Finansdepartementet vurdere IFRS for SMEs*. <https://www.regnskapsstiftelsen.no/nrs-ber-finansdepartementet-vurdere-ifrs-for-smes/>
- NOU 1995: 30. (1995). *Ny regnskapslov*. Finansdepartementet.  
<https://www.regjeringen.no/no/dokumenter/nou-1995-30/id115910/>
- Peecher, M. E. & Solomon, I. (2001). Theory and Experimentation in Studies of Audit Judgments and Decisions: Avoiding Common Research Traps. *International Journal of Auditing*, 5(3), 193-203. <https://doi.org/10.1111/1099-1123.00335>
- Pettersen, L. I. (2009). Fortsatt drift-forutsetningen. *Revisjon og regnskap*, (1), 21-25.  
<https://www.revregn.no/asset/pdf/2009/1-21-5.pdf>
- Preston, C. C. & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104(1), 1-15. [https://doi.org/10.1016/S0001-6918\(99\)00050-5](https://doi.org/10.1016/S0001-6918(99)00050-5)
- Prop. 42 (1997-1998). *Om lov om årsregnskap m.v. (regnskapsloven)*. Finansdepartementet.  
<https://www.regjeringen.no/no/dokumenter/otprp-nr-42-1997-98-/id120622/>
- PCAOB. (2019). *Strategic plan 2020-2024*. [https://pcaob-assets.azureedge.net/pcaob-dev/docs/default-source/about/administration/documents/strategic\\_plans/strategic-plan-2020-2024.pdf?sfvrsn=776073d3\\_4](https://pcaob-assets.azureedge.net/pcaob-dev/docs/default-source/about/administration/documents/strategic_plans/strategic-plan-2020-2024.pdf?sfvrsn=776073d3_4)

- 
- Puthukulam, G., Ravikumar, A., Sharma, R. V. K. & Meesaala, K. M. (2021). Auditors' Perception on the Impact of Artificial Intelligence on Professional Skepticism and Judgment in Oman. *Universal Journal of Accounting and Finance*, 9(5), 1184-1190. <https://doi.org/10.13189/ujaf.2021.090527>
- PwC. (2017). *Harnessing the power of AI to transform the detection of fraud and error*. <https://www.pwc.com/gx/en/about/stories-from-across-the-world/harnessing-the-power-of-ai-to-transform-the-detection-of-fraud-and-error.html>
- PwC. (2020). *Effektiv revisjon med de beste verktøyene*. <https://www.pwc.no/no/tjenester/revisjon/innovasjon-og-verifikasjonstjenester.html>
- PwC. (u.å.a). *Hva er Big Data?* <https://www.pwc.no/no/teknologi-omstilling/digitalisering-pa-1-2-3/big-data---data-lake.html>
- PwC. (u.å.b). *Hva er kunstig intelligens?* <https://www.pwc.no/no/teknologi-omstilling/digitalisering-pa-1-2-3/kunstig-intelligens.html>
- Raman, K., Svore, K. M., Gilad-Bachrach, R. & Burgres, C. J. (2012). Learning from mistakes: towards a correctable learning algorithm. *Association for Computing Machinery*, 1930-1934. <https://doi.org/10.1145/2396761.2398546>
- Rapoport, M. (2016). Auditors count on tech for backup. *Wall Street Journal*. <https://www.wsj.com/articles/auditing-firms-count-on-technology-for-backup-1457398380>
- Regjeringen. (u.å.). *Kompensasjonsordning for næringslivet*. <https://www.kompensasjonsordning.no>
- Regnskapsloven. (1998). *Lov om årsregnskap m.v.* (LOV-1998-07-17-56). Lovdata. <https://lovdata.no/dokument/NL/lov/1998-07-17-56>
- Revisorloven. (2020). *Lov om revisjon og revisorer*. (LOV-2020-11-20-128). Lovdata. <https://lovdata.no/dokument/LTI/lov/2020-11-20-128>
- Rieland, R. (2018). *Artificial Intelligence Is Now Used to Predict Crime. But Is It Biased?* Smithsonian Magazine. <https://www.smithsonianmag.com/innovation/artificial-intelligence-is-now-used-predict-crime-is-it-biased-180968337/>
- Rousseau, D., Sitkin, S. B., Burt, R. & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, 23(3), 393-404. <https://doi.org/10.5465/amr.1998.926617>
- Russel, S. & Norvig, P. (2016). *Artificial Intelligence: A Modern Approach*. Boston: Pearson Education Ltd.



- 
- Rødssæteren, K. & Gøbel, A. (2010). Revisjon av forutsetningen om fortsatt drift. *Praktisk økonomi & finans*, 25(4), 13-25.  
<https://doi.org/10.18261/ISSN1504-2871-2009-04-03>
- Samuel, A. L. (1952). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, 3(3), 210-229.  
<https://doi.org/10.1147/rd.33.0210>
- Saunders, M., Lewis, P. & Thornhill, A. (2016). Research Methods for Business Students (7. Utgave). *Pearson Education Limited*.
- Sjåstad, H. (2019). *Algoritme-aversjon*. Magma. <https://old.magma.no/algoritme-aversjon>
- Skatteetaten. (u.å.). *Tiltakspakker i forbindelse med korona-situasjonen*. Hentet 25. mars 2022 fra <https://www.skatteetaten.no/tiltakspakker/>
- Spector, P. E. (2020). Mastering the Use of Control Variables: the Hierarchical Iterative Control (HIC) Approach. *Journal of Business and Psychology*, 36, 737-750.  
<https://doi.org/10.1007/s10869-020-09709-0>
- Sternthal, B., Dholakia, R. & Leavitt, C. (1978). The persuasive effect of source credibility: Tests of cognitive response. *Journal of Consumer Research*, 4(4), 252–260.  
<https://doi.org/10.1086/208704>
- Store norske leksikon. (2018). *Enron Corporation. Internasjonal teknologi- og industrihistorie*. [https://snl.no/Enron\\_Corporation](https://snl.no/Enron_Corporation)
- Store norske leksikon. (2019a). *Deduksjon. Fremmedord med latinsk og gresk opphav*.  
<https://snl.no/deduksjon>
- Store norske leksikon. (2019b). *WorldCom. Mediekonsern*. <https://snl.no/WorldCom>
- Store norske leksikon. (2020). *Arthur Andersen. Bedrifter i internasjonal næringslivshistorie*. [https://snl.no/Arthur\\_Andersen](https://snl.no/Arthur_Andersen)
- Teknologirådet. (2018). *Kunstig intelligens – muligheter, utfordringer og en plan for Norge*.  
<https://teknologiradet.no/wp-content/uploads/sites/105/2018/09/Rapport-Kunstig-intelligens-og-maskinlaering-til-nett.pdf>
- Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L., Robson, R., Thabane, M., Giangregorio, L. & Goldsmith, C. (2010). A tutorial on pilot studies: the what, why and how. *BMC Med Res Methodol*, 10(1), 1-10. <https://doi.org/10.1186/1471-2288-10-1>
- Tidemann, A. (u.å.). *Nevralt nettverk*. Store norske leksikon. Hentet 29. mars 2022 fra [https://snl.no/nevralt\\_netverk](https://snl.no/nevralt_netverk)



- Tidemann, A. & Elster, A. C. (2022, 18. januar). Maskinl ring. SNL. <https://snl.no/maskinl%C3%A6ring>
- Tran y, K. E. (2020). *hypotese*. SNL. <https://snl.no/hypotese>
- Tran y, K. E. (2021, 2. desember). *Induksjon*. SNL. [https://snl.no/induksjon\\_-\\_filosofi](https://snl.no/induksjon_-_filosofi)
- Tripp, D. (1993). *Critical Incidents in Teaching: Developing professional judgement*. Routledge. <https://doi.org/10.4324/9780203802014>
- Trochim, W. M., Donnelly, J. P. & Arora, K. (2016). *Research methods: The essential knowledge base*. Cengage Learning.
- Trotman, K. T. (2001). Design issues in Audit JDM Experiments. *International Journal of Auditing*, 5, 181-192. <https://doi.org/10.1111/1099-1123.00334>
- Vereschak, O., Bailly, G. & Caramiaux, B. (2021). How to Evaluate Trust in AI-Assisted Decision Making? A Survey of Empirical Methodologies. *Proceedings of the ACM on Human-Computer Interaction*, 5, 1-39. <https://doi.org/10.1145/3476068>
- Yin, M., Vaughan, J. W. & Wallach, H. (2019). Understanding the Effect of Accuracy on Trust in Machine Learning Models. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1-12. <https://doi.org/10.1145/3290605.3300509>
- Yoon, K., Hoogduin, L. & Zhang, L. (2015). Big Data as Complementary Audit Evidence. *Accounting Horizons*, 29(2), 431–438. <https://doi.org/10.2308/acch-51076>
- Zednik, C. (2019). Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence. *Philosophy & Technology*, 34, 265-288. <https://doi.org/10.1007/s13347-019-00382-7>
- Zhang, Y., Liao, Q. V. & Bellamy, R. K. E. (2020). Effect of Confidence and Explanation on Accuracy and Trust Calibration in AI-Assisted Decision Making. *Association for Computing Machinery*, 295-305. <https://doi.org/10.1145/3351095.3372852>

## Vedlegg

### Vedlegg 1: Eksperimentelt case og undersøkelse

---

Hei!

Takk for at du tar deg tid til å delta i undersøkelsen i forbindelse med vår masterutredning. Estimert tidsbruk er 7 minutter. Vi ber vennlig om at informasjonen leses nøye og at du svarer ærlig og selvstendig.

Undersøkelsen er anonym.

---

#### **Din rolle**

Du er revisor for et flyselskap og er i ferd med å revidere årsregnskapet for 2021. I den forbindelse skal du vurdere hvorvidt det foreligger hendelser eller forhold som kan skape tvil av betydning om enhetens evne til fortsatt drift.

#### **Informasjon om flyselskapet**

Flyselskapet driver innen markedet for interkontinental godsfrakt. Selskapet er en mindre aktør i markedet med sine 35 fraktfly, 25 av disse går på lengre kontrakter og ti i spotmarkedet. De to siste årene har markedet vært tilnærmet normalt. Flyselskapet har imidlertid vært under press fra de større aktørene i markedet og har derfor en noe svak balanse. Blant annet er gjeldsgraden svakere enn bransjeoptimalt, og kontantbeholdningen lavere enn ønsket fra ledelsen sin side.

Etter årets to første måneder er inntjeningen i selskapet lavere enn budsjettet, blant annet som følge av høyere energipriser. En betydelig del av flyflåten består av aldrende fly, som er mindre energieffektive enn dagens gjennomsnittlige bransjestandard. Vedvarende høye energipriser reduserer konkurranseevnen til selskapet, sett i forhold til hovedkonkurrentene. Positive driftsmarginer avhenger av at energiprisene reduseres gjennom andre halvår inneværende år.

Nøkkeltallet kortsiktige fordringer/kortsiktig gjeld er svakt i kvartalsregnskapet for Q4 2021, som i hovedsak skyldes forfall av hovedstol på et obligasjonslån pålydende 150 MUSD om syv måneder. Selskapets ledelse har over lengre tid vært i forhandlinger om å restrukturere den aktuelle gjelden, men har per nå ikke undertegnet ny låneavtale. Ledelsen forhandler med kreditorer som de allerede har et forhold til. Refinansiering av gjelden er helt nødvendig, ettersom selskapet ikke har tilstrekkelige likvider til å nedbetale gjelden på forfallstidspunktet.

Fremtidsutsiktene til selskapets ledelse tyder på økte driftsmarginer i kommende periode som følge av en begrenset tilbudsside. Den totale flyflåten i markedet ventes å øke minimalt i perioden fremover, på grunn av utfordringer hos flyprodusentene. Selskapet har opsjoner på å kjøpe fem nye fly med leveranse i slutten av året. Utøvelse av opsjonene avhenger av at selskapet får solgt flere av de eldste flyene i flåten. Markedet for denne type fly er imidlertid preget av nye internasjonale miljøkrav, og salgsprisen er derfor usikker.

Flyselskapet er nylig underrettet om et mulig søksmål som følge av manglende etterlevelse av sikkerhetsrutiner, av internasjonale luftfartsmyndigheter. Myndighetene ser svært alvorlig på saken, og har historisk uttrykt dette gjennom betydelige økonomiske sanksjoner.

***Gruppe 1 får ikke noe mer informasjon utover det hittil ovennevnte***

***Gruppe 2 får følgende ytterligere informasjon:***

#### **Hjelpeverktøy**

Ditt revisjonsselskap har nylig valgt å ta i bruk et kunstig intelligens-verktøy som skal vurdere hvorvidt regnskapet kan utarbeides under forutsetningen om fortsatt drift.

**KI-verktøyets vurdering:** KI-verktøyet vurderer at regnskapet kan utarbeides under forutsetning om fortsatt drift.

***Gruppe 4 får følgende ytterligere informasjon, utover informasjonen som gis gruppe 1:***

#### **Hjelpeverktøy**

Revisjonsbransjen har de siste årene stått overfor store digitale omveltninger, og har tatt i bruk en rekke digitale analyseverktøy. Ditt revisjonsselskap har nylig valgt å ta i bruk et kunstig intelligens-verktøy som skal vurdere hvorvidt regnskapet kan utarbeides under forutsetningen

om fortsatt drift. Verktøyet er bygget på bruk av en algoritme utviklet av Microsoft.

Algoritmen har vist seg å ha svært høy prediksjonsevne, med 91% treffsikkerhet (mot revisors 70%). Datasettet lagt til grunn ved utarbeidelse av algoritmen, består av 4370 årsregnskap fra perioden 2011-2020. KI-verktøyet er i stand til å prosessere store mengder data umiddelbart, og fanger opp detaljer i datamengden i større grad enn mennesker. Algoritmen er i stand til å foreta en mer objektiv vurdering enn mennesker, ettersom blant annet relasjoner og interessekonflikter ikke er inkludert i input-dataene.

**KI-verktøyet vurdering:** KI-verktøyet vurderer at regnskapet kan utarbeides under forutsetning om fortsatt drift.

*Spørsmålene som angis under ble gitt alle gruppene, med mindre annet er angitt i parentes.*

**Vi ber vennlig om at du svarer på følgende spørsmål:**

	Liten usikkerhet	2	3	4	5	Svært stor usikkerhet
	1					6
Hvordan vurderer du usikkerheten til enhetens evne til fortsatt drift?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Lite sannsynlig	2	3	4	5	Svært sannsynlig
	1					6
Hvor sannsynlig er det at du vil utføre ytterligere handlinger før du konkluderer på at forutsetningen om fortsatt drift kan legges til grunn?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hvor sannsynlig, gitt nåværende informasjon, er det at du vil legge til et avsnitt om usikkerhet knyttet til fortsatt drift i revisjonsberetningen?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Manipulasjonssjekk 1:** Fikk du informasjon om at et KI-verktøy ble tatt i bruk?

- Ja
- Nei

**Manipulasjonssjekk 2 (gruppe 2 og 4):** Fikk du informasjon om prediksjonsevnen til KI-verktøyet?

- Ja
- Nei

**Kontrollspørsmål (gruppe 2 og 4):**

	I svært liten grad / svært lav				I svært stor grad / svært høy	
	1	2	3	4	5	6
I hvor stor grad vektla du vurderingen til AI i din egen vurdering?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hvor høy tillit har du til algoritmen?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Kjønn?**

- Mann
- Kvinne
- Annet

**Alder?**

- Under 20
- 20-25
- 26-30
- Over 30

**Har du arbeidserfaring innen revisjon?**

- Nei
- Mindre enn ett år
- 1-2 år
- 2-3 år
- 3-4 år
- Mer enn fire år

**Hvilken masterretning går du/har du gått?**

- Master i regnskap og revisjon
- Master i økonomi og administrasjon
- Andre

## Vedlegg 2: Cronbach's Alpha

### Indeks profesjonelt skjønn

#### Reliability Statistics

Cronbach's Alpha	N of Items
.833	3

### Indeks algoritme-aversjon

#### Reliability Statistics

Cronbach's Alpha	N of Items
.743	2

## Vedlegg 3: Shapiro-Wilk test

Tests of Normality							
	Gruppe	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Usikkerhet	1	.200	17	.068	.916	17	.128
	2	.363	11	<.001	.810	11	.013
	4	.292	13	.003	.896	13	.118
Handling	1	.359	17	<.001	.715	17	<.001
	2	.353	11	<.001	.649	11	<.001
	4	.187	13	.200*	.883	13	.078
Beretning	1	.161	17	.200*	.915	17	.123
	2	.191	11	.200*	.863	11	.064
	4	.153	13	.200*	.914	13	.207
Indeks.ps	1	.160	17	.200*	.935	17	.259
	2	.297	11	.008	.797	11	.009
	4	.185	13	.200*	.936	13	.404

\*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

## Tests of Normality

	Gruppe	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Vektla	2.00	.310	11	.004	.866	11	.069
	4.00	.237	13	.044	.844	13	.024
Tillit	2.00	.219	11	.146	.889	11	.134
	4.00	.234	13	.049	.885	13	.084
Indeks.	2.00	.142	11	.200*	.964	11	.820
	4.00	.222	13	.081	.929	13	.327

\*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

## Vedlegg 4: Hypotese 1

### Independent Samples T-test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
		F	Sig.	t	df	Significance		Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
						One-Sided p	Two-Sided p			Lower	Upper
Usikkerhet	Equal variances assumed	.023	.879	.991	28	.165	.330	.430	.434	-.459	1.318
	Equal variances not assumed			.962	22.579	.173	.346	.430	.447	-.496	1.356
Handling	Equal variances assumed	5.995	.021	2.298	28	.015	.029	1.027	.447	.112	1.943
	Equal variances not assumed			2.113	16.469	.025	.050	1.027	.486	-.001	2.056
Beretning	Equal variances assumed	3.020	.093	1.187	28	.123	.245	.643	.541	-.466	1.751
	Equal variances not assumed			1.133	20.638	.135	.270	.643	.567	-.538	1.823
Indeks 1	Equal variances assumed	1.569	.221	1.637	28	.056	.113	.699849170	.427578723	-.176006140	1.57570448
	Equal variances not assumed			1.543	19.157	.070	.139	.699849170	.453593945	-.249006312	1.64870465

### Mann-Whitney Test

#### Test Statistics<sup>a</sup>

	Handling
Mann-Whitney U	67.500
Wilcoxon W	158.500
Z	-1.920
Asymp. Sig. (2-tailed)	.055
Exact Sig. [2*(1-tailed Sig.)]	.072 <sup>b</sup>
Exact Sig. (2-tailed)	.060
Exact Sig. (1-tailed)	.031
Point Probability	.006

a. Grouping Variable: Gruppe

b. Not corrected for ties.

## Vedlegg 5: Hypotese 2

### Independent Samples T-test

		Levene's Test for Equality of Variances				t-test for Equality of Means				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
Usikkerhet	Equal variances assumed	.950	.340	.769	22	.225	.450	.350	.455	-.593	1.293
	Equal variances not assumed			.801	19.958	.216	.432	.350	.436	-.560	1.260
Handling	Equal variances assumed	8.132	.009	2.286	22	.016	.032	1.161	.508	.108	2.214
	Equal variances not assumed			2.453	14.896	.013	.027	1.161	.473	.151	2.170
Beretning	Equal variances assumed	2.359	.139	1.098	22	.142	.284	.685	.624	-.609	1.979
	Equal variances not assumed			1.132	21.313	.135	.270	.685	.605	-.573	1.943
Indeks 1	Equal variances assumed	5.357	.030	1.606	22	.061	.123	.731934732	.455807009	-.213351149	1.67722061
	Equal variances not assumed			1.717	15.565	.053	.106	.731934732	.426355131	-.173956765	1.63782623

### Mann-Whitney Test

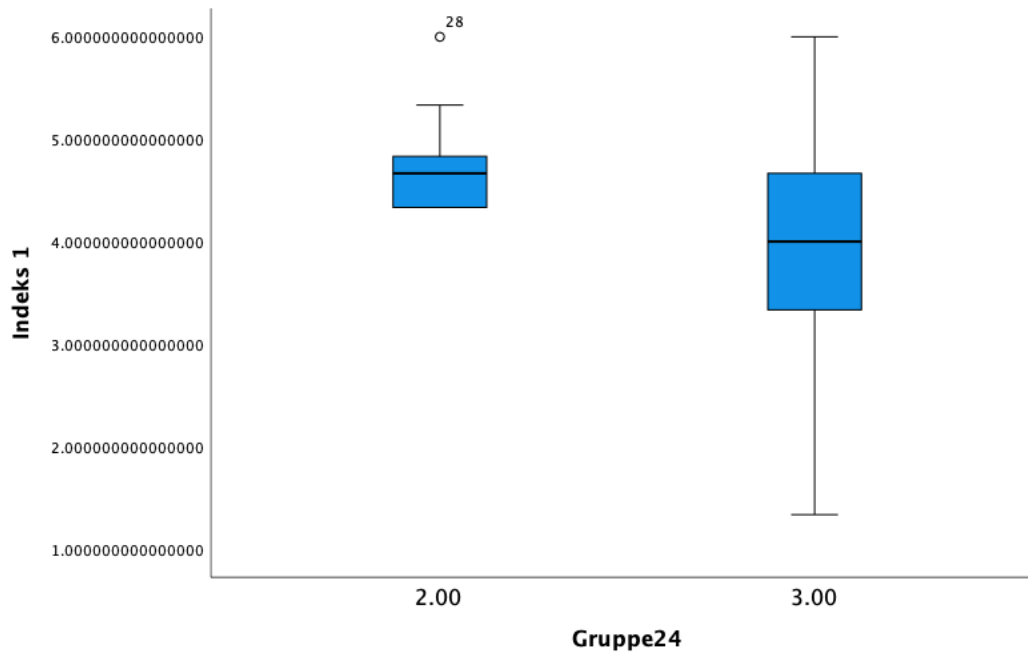
Test Statistics <sup>a</sup>			
	Usikkerhet	Handling	Indeks 1
Mann-Whitney U	65.000	39.500	44.500
Wilcoxon W	156.000	130.500	135.500
Z	-.413	-1.965	-1.585
Asymp. Sig. (2-tailed)	.680	.049	.113
Exact Sig. [2*(1-tailed Sig.)]	.733 <sup>b</sup>	.063 <sup>b</sup>	.119 <sup>b</sup>
Exact Sig. (2-tailed)	.695	.051	.118
Exact Sig. (1-tailed)	.361	.028	.059
Point Probability	.048	.005	.004

a. Grouping Variable: Gruppe

b. Not corrected for ties.



## Boksplott



Quade Nonparametric ANCOVA: Indeks profesjonelt skjønn

### Quade Nonparametric Analysis of Covariance

F	DFH	DFE	P Value
1.070	1	22	.312

### Pairwise Comparisons of Groups

Comparison	t	DF	P Value
2.00 vs. 3.00	1.035	22	.312

## Quade Nonparametric ANCOVA: «Ytterligere handlinger»

### Quade Nonparametric Analysis of Covariance

F	DFH	DFE	P Value
1.923	1	22	.179

### Pairwise Comparisons of Groups

Comparison	t	DF	P Value
2.00 vs. 3.00	1.387	22	.179

## Vedlegg 6: Hypotese 3

### Independent Samples T-test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
		F	Sig.	t	df	Significance		Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
						One-Sided p	Two-Sided p			Lower	Upper
Vektla	Equal variances assumed	.088	.769	-1.207	22	.120	.240	-.441	.365	-1.198	.317
	Equal variances not assumed			-1.224	21.998	.117	.234	-.441	.360	-1.187	.306
Tillit	Equal variances assumed	.049	.827	-1.737	22	.048	.096	-.734	.423	-1.611	.142
	Equal variances not assumed			-1.743	21.605	.048	.095	-.734	.421	-1.609	.140
Indeks 2	Equal variances assumed	.006	.939	-1.684	22	.053	.106	-.5874	.3488	-1.3107	.1359
	Equal variances not assumed			-1.686	21.445	.053	.106	-.5874	.3483	-1.3109	.1361

### Mann-Whitney Test

#### Test Statistics<sup>a</sup>

	Vektla
Mann-Whitney U	50.000
Wilcoxon W	116.000
Z	-1.328
Asymp. Sig. (2-tailed)	.184
Exact Sig. [2*(1-tailed Sig.)]	.228 <sup>b</sup>
Exact Sig. (2-tailed)	.200
Exact Sig. (1-tailed)	.104
Point Probability	.022

a. Grouping Variable: Gruppe

b. Not corrected for ties.