

Stig Langlo Brøndbo
Line Flemmos Larsen

Konkursprediksjon med ubehandlede regnskapsdata

– En kvantitativ studie av prediksjonevnen til
nevrale nettverk basert på data fra norske
aksjeselskaper

Masteroppgave i Økonomi og administrasjon

Veileder: Arild Brandrud Næss

Mai 2022

Stig Langlo Brøndbo
Line Flemmos Larsen

Konkursprediksjon med ubehandlede regnskapsdata

– En kvantitativ studie av prediksjonsevnen til nevralt nettverk basert på data fra norske aksjeselskaper

Masteroppgave i Økonomi og administrasjon
Veileder: Arild Brandrud Næss
Mai 2022

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
NTNU Handelshøyskolen



NTNU

Kunnskap for en bedre verden

Forord

Denne oppgaven er skrevet som en avsluttende del av vår mastergrad innen økonomi og administrasjon ved NTNU Handelshøyskolen. Oppgaven ble skrevet våren 2022 innenfor hovedprofilen *business analytics* og utgjør 30 studiepoeng.

Vi vil rette en stor takk til veilederen vår Arild Brandrud Næss for gode diskusjoner og god oppfølging. Vi vil også takke Ranik Raaen Wahlstrøm som har hjulpet oss med tilgang til og tilpasning av datagrunnlaget som er benyttet.

Vi vil videre takke familie, samboere og medstudenter for god støtte, innspill og motiverende ord underveis i prosessen. Det har vært en lærerik prosess med både opp- og nedturer underveis.

Innholdet i denne oppgaven står for forfatterens regning.

Sammendrag

Temaet i denne oppgaven er konkursprediksjon. Formålet er å undersøke hvordan valg av uavhengige variabler kan påvirke prediksjonsevnen til nevrale nettverk innenfor konkursprediksjon. I denne oppgaven blir det undersøkt hvorvidt bruk av ubehandlede regnskapsdata kan gi bedre prediksjoner enn variabelsett som er benyttet i tidligere forskning. Det finnes ingen universell konsensus på hvilke uavhengige variabler som bør brukes til konkursprediksjon, men det er flere variabler som går igjen i den tidligere forskningen på området. Oppgaven presenterer fire ulike variabelsett bestående av ubehandlede regnskapsdata hvor to av disse også inneholder finansielle nøkkeltall fra tidligere forskning. Variabelsett som er basert på Altmans Z-score modell og Norges Banks SEBRA-modell brukes som sammenlikning for disse fire variabelsettene. Det settes opp modeller med nevrale nettverk for hvert variabelsett og disse vurderes opp mot hverandre. Det lages også referansemodeller med logistisk regresjon for hvert variabelsett. Datagrunnlaget som benyttes i denne oppgaven er en samling av alle norske ikke-finansielle årsregnskaper i perioden 2006–2019, og inneholder til sammen 3 876 923 ikke-konsoliderte norske årsregnskaper.

Problemstillingen i denne oppgaven er som følger:

I hvilken grad kan ubehandlede regnskapsdata forbedre konkursprediksjon med dyplæringsmodeller?

Oppgaven konkluderer med at det er nyttig å bruke ubehandlede regnskapsdata i nevrale nettverk, siden resultatene tyder på at finansielle nøkkeltall i slike modeller kan erstattes av regnskapstall. Dette begrunnes med at slike modeller med dagens datakraft er i stand til å finne sammenhengene i regnskapstallene som disse finansielle nøkkeltallene er skapt for å illustrere.

Abstract

The subject of this thesis is bankruptcy prediction. The purpose is to explore how the choice of independent variables can affect the predictive power of neural networks within bankruptcy prediction. The thesis examines whether the use of unprocessed accounting data can provide better predictions than variable sets used in previous research. There is no universal consensus on which independent variables should be used for bankruptcy prediction, but there are several variables that recur in relevant research. The thesis presents four different models constructed with unprocessed accounting data, two of which contain financial ratios found in earlier research. Altman's Z-score model and Norges Bank's SEBRA model are used as a comparison for these four models. Models with neural networks are set up for each variable set and evaluated against each other. Reference models are created with logistic regression for each variable set. The data used in this thesis is a collection of all Norwegian 3 876 923 non-financial annual accountings in the period 2006–2019.

The research question is the following:

To what extent can unprocessed accounting data improve bankruptcy prediction with deep learning models?

The thesis concludes that unprocessed accounting data can replace financial ratios in neural networks. Models with sufficient computing power are found to make connections in accounting items similar to those of financial ratios.

Innholdsfortegnelse

1	Innledning	1
1.1	Motivasjon	1
1.2	Dyplæring	3
1.3	Problemstilling.....	4
1.4	Oppgavens struktur	5
2	Tidligere forskning	6
2.1	Altman-modellene	6
2.2	SEBRA-modellene.....	7
2.3	Andre metoder for konkursprediksjon.....	8
2.4	Design av nevrale nettverksmodeller i tidligere forskning.....	9
2.5	Variabler i tidligere forskning.....	10
2.6	Kategorier av nøkkeltall.....	11
3	Data.....	12
3.1	Beskrivelse og behandling av datagrunnlag	12
3.2	Avhengig variabel.....	13
3.3	Uavhengige variabler.....	15
3.3.1	Variabelsett Altman	16
3.3.2	Variabelsett SEBRA	16
3.3.3	Variabelsett Regnskap	18
3.3.4	Variabelsett Regnskap-	18
3.3.5	Variabelsett Regnskap+Altman	18
3.3.6	Variabelsett Regnskap+Altman-	19
3.4	Utarbeidelse av datasett	19
3.4.1	Dele på 0	19
3.4.2	Winsorizing.....	19
3.4.3	Trenings-, validerings- og testsett.....	20
3.4.4	Standardisering	21
3.5	Datakvalitet.....	21
4	Metode	23
4.1	Balansering av datasettet	23
4.2	Estimering.....	25
4.2.1	Hyperparameteroptimering for nevrale nettverk	25
4.2.2	Oppsett av modeller	28
4.2.3	Tidshorisont for konkurs.....	29
4.3	Evaluering.....	30
4.4	Repliserbarhet.....	33
5	Resultater	34
5.1	Resultater fra logistisk regresjon	34

5.2	Resultater fra nevrale nettverk	34
5.3	SHAP-verdier.....	35
5.4	McNemars test	38
6	Drøfting	40
6.1	Modeller med ubehandlede regnskapsdata og modeller med nøkkeltall	40
6.2	Modeller med SEBRA og ALTMAN	41
6.3	Sammenlikning av de fire variabelsettene med ubehandlede regnskapsdata	42
6.4	Modeller med Regnskap og Regnskap-	42
6.5	Modeller med Regnskap+Altman og Regnskap+Altman-	44
6.6	Konsekvenser for banker	45
6.7	Kostnader ved ulike typer feil.....	45
6.8	Modellenes kompleksitet	46
7	Konklusjon.....	47
7.1	Konkluderende bemerkninger.....	47
7.2	Utfordringer og begrensninger.....	48
7.3	Forslag til videre forskning.....	49
	Referanseliste	50
	Appendiks.....	53
	A1: Variabler i variabelsett.....	54
	A2: Utarbeidelse av Altman-nøkkeltall	57
	A3: <i>Confusion matrix</i> for modellene	58
	A4: ROC-diagrammer for modellene	59
	A5: SHAP-analyser av modeller med nevrale nettverk.....	61
	A6: Datamaskinspesifikasjoner og Python-bibliotek	63
	A7: Treffprosent og tap på trenings- og valideringssett for nevrale nettverk.....	64

Figurliste

Figur 1: Konkurser per år i perioden 2006 til 2021	2
Figur 2: Illustrasjon av et dypt nevralt nettverk.....	3
Figur 3: Andel av totale konkursen i datagrunnlaget basert på selskapsalder.....	15
Figur 4: Fordeling av konkursen før og etter undersampling.....	24
Figur 5: Illustrasjon av de anvendte aktiveringsfunksjonene	27
Figur 6: Eksempel på ROC-diagram.....	32
Figur 7: SHAP-analyse av variabelsett Altman.....	36
Figur 8: SHAP-analyse av variabelsett SEBRA	37
Figur 9: SHAP-analyse av variabelsett Regnskap.....	37
Figur 10: SHAP-analyse av variabelsett Regnskap+Altman	38

Tabelliste

Tabell 1: Fordeling av konkursen per år i datagrunnlaget i perioden 2006-2017	14
Tabell 2: Variabler i variabelsett Altman	16
Tabell 3: Variabler i variabelsett SEBRA.....	17
Tabell 4: Verdier fra hyperparameteroptimering	28
Tabell 5: Confusion matrix med oversikt over ulike typer feil.....	30
Tabell 6: Oppsummering av resultater fra logistiske regresjoner	34
Tabell 7: Oppsummering av resultater fra nevrale nettverk	35
Tabell 8: p-verdier fra McNemars test.....	39

1 Innledning

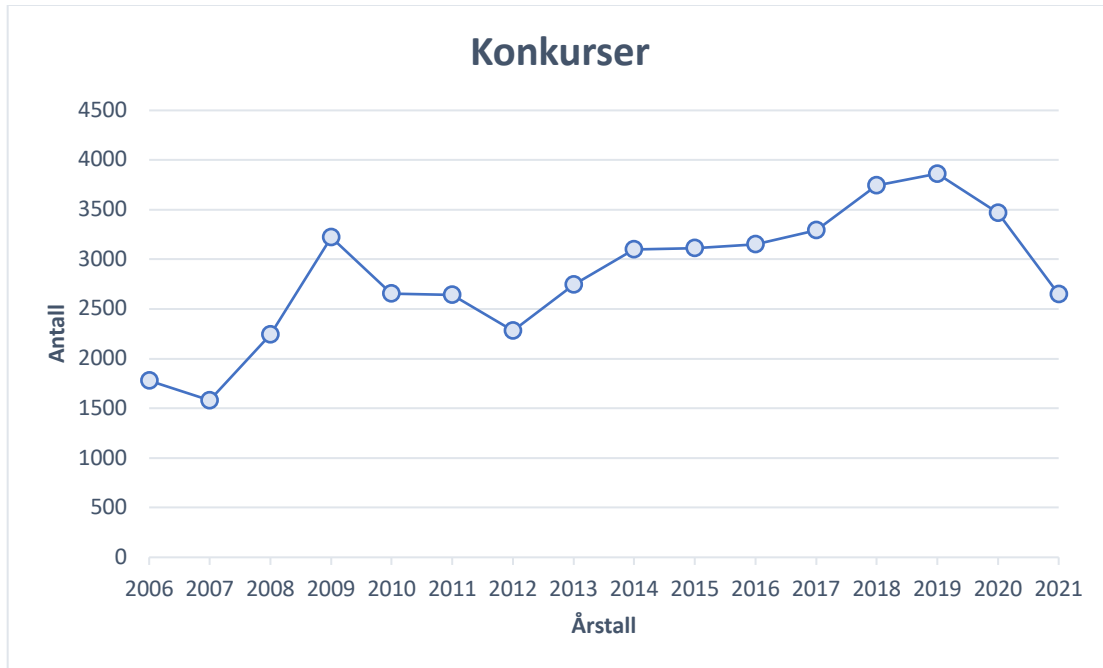
1.1 Motivasjon

«Norske banker har historisk tapt vesentlig mer på utlån til norske ikke-finansielle foretak enn på utlån til husholdninger» (Finanstilsynet, 2021, s. 2).

Utlån vil alltid føre med seg en viss risiko. Som det kommer frem av sitatet ovenfor, utgjør utlån til selskaper en stor andel av utlånstapet til norske banker. For banker er utlån til selskaper særlig avhengig av selskapets risiko for konkurs. Selv om banker er underlagt strenge regler for utlån, vil de likevel måtte vurdere selv om en lånesøknad er fornuftig å innvilge. Dette fremhever viktigheten av å gjøre gode analyser av selskapene som søker om lån. Når selskaper misligholder gjeld kan det føre til store tap for banken, men det kan også få store ringvirkninger blant både ansatte og andre interessenter.

En konkurs i et selskap er en lengre prosess som starter med at en part begjærer selskapet konkurs eller at selskapet selv melder oppbud. Konkursloven (1986, § 60) sier følgende: «Er skyldneren insolvent, skal vedkommendes bo tas under konkursbehandling når det begjæres av skyldneren eller en fordringshaver.» Vilkårene for at noen skal være insolvent, og dermed kunne bli begjært konkurs, er at de må være både illikvide og insuffisiente (Konkursrådet, 2012). For et selskap vil det si at de ikke har nok midler, og heller ikke andre omsettelige verdier, til å dekke sine løpende forpliktelser (Konkursloven, 1986, § 61). Konkursbehandlingen kan ta opptil flere år avhengig av kompleksiteten i selskapet, samt om behandlingen er avhengig av eksterne prosesser som for eksempel andre konkursbehandlinger.

Det er mange unge selskaper som går konkurs (Statistisk sentralbyrå, 2021), og det at antall konkurser øker år for år er derfor logisk siden antall nyetableringer også øker (Statistisk sentralbyrå, 2021). I figur 1 vises det at det er en økning i antall konkurser i årene 2008 og 2009 (Statistisk sentralbyrå, 2022). Dette er en naturlig ettervirkning av finanskrisen i 2008. I figuren kan man også se at årene 2020 og 2021 har en nedgang i antall konkurser. Dette skyldes trolig ekstraordinære tiltak og statlige kompensasjonsordninger som har blitt innført i forbindelse med covid-19. Disse tiltakene kan ha hjulpet selskaper som sto i fare for å gå konkurs. Ettersom denne oppgaven ikke vektlegger tidstrenden, som beskrevet i seksjon 3.2, antas det at disse og liknende hendelser ikke vil å få store konsekvenser for oppgaven. Siden datasettet strekker seg over relativt få år blir det heller ikke tatt høyde for inflasjon i oppgaven, ettersom dette antas å ha liten innvirkning på resultatene.



Figur 1: Konkurser per år i perioden 2006 til 2021

Når en bank skal vurdere en lånesøknad er det viktig for banken å vurdere sannsynligheten for mislighold av lånet. For å gjøre denne vurderingen på en systematisk og objektiv måte er det mulig å lage modeller. Det å kunne lage en modell for å predikere om et selskap vil gå konkurs eller ikke vil kunne gi større forutsigbarhet for banker, noe som gjør at de kan planlegge på lenger sikt. Bankene kan benytte prediksjonene til å avgjøre hvor mye de skal låne ut til ulike selskaper, og dermed minimere utlånstapet. Konkursprediksjon er derfor en av de viktigste beslutningsrelevante aktivitetene som finnes (Kirkos, 2015).

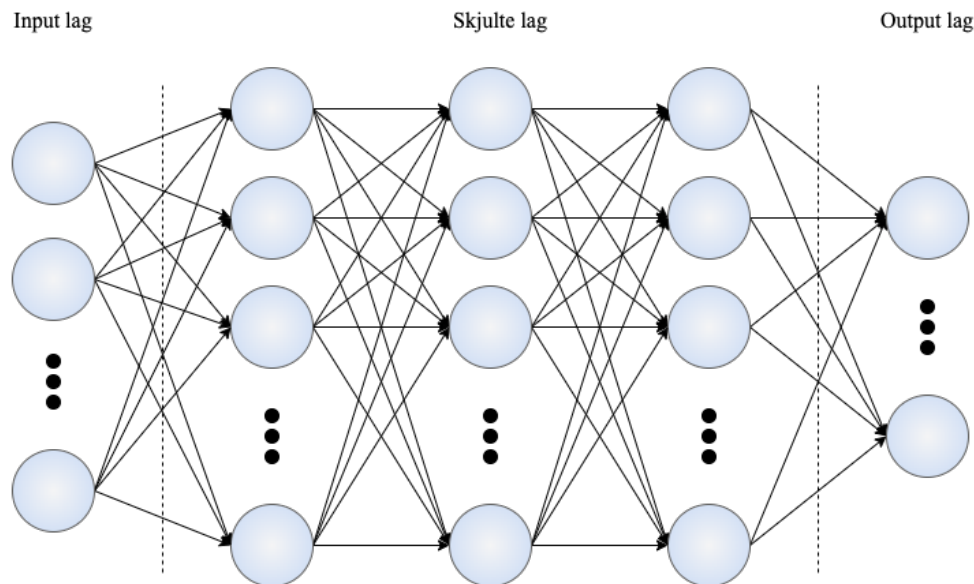
Nytten av å forbedre konkursprediksjonsmodeller er høy, da selv små forbedringer kan medføre stor økonomisk gevinst (Stein, 2005). Modeller som kan benyttes til denne typen prediksjon er derfor i konstant utvikling innenfor fagfeltet, og særlig i finansielle foretak som har interesse av denne typen modeller. Ved å finne mønstre basert på regnskapstall kan man si noe om sannsynligheten for at et selskap vil gå konkurs eller ikke. Formålet med denne oppgaven er derfor å se hvordan valg av uavhengige variabler påvirker dyplæringsmodeller, og om bruk av ubehandlede regnskapsdata kan redusere eller fjerne behovet for å sette sammen finansielle nøkkeltall for å predikere konkurs.

Banker sitter på mye data knyttet til konkurser, og er derfor avhengig av modeller som er gode på å behandle store mengder data for å kunne drive med konkursprediksjon. Det finnes blant annet metoder innen maskinlæring som har stort potensiale dersom de brukes til dette formålet. Neste seksjon vil gi en innføring i nevrale nettverk med fokus på dyplæring for å etablere noen sentrale begreper og en kontekst for temaet i oppgaven.

1.2 Dyplæring

Maskinlæring er betegnelsen på at en datamaskin setter opp og forbedrer modeller ved å finne mønstre i den tilgjengelige datamengden (Russel & Norvig, 2021, s. 669). Kunstige nevralt nettverk er betegnelsen på en type ikke-lineære maskinlæringsmodeller (Paraschiv et al., 2021), hvor modellene består av komplekse algebraiske kretser med justerbar forbindelsesstyrke (Russel & Norvig, 2021, s. 801). Det er en generell oppfatning om at de nevralt nettverkene er inspirert av hvordan det menneskelige nervesystemet fungerer (Ravi Kumar & Ravi, 2007). Chollet (2021, s. 7) mener imidlertid at dette er en misledende sammenlikning og hevder at nevralt nettverk først og fremst bør sees på som et matematisk rammeverk som forsøker å finne sammenhenger i data.

De nevralt nettverksmodellene har et inputlag, et eller flere skjulte lag, og et outputlag. En fremstilling av et nevralt nettverk vises i figur 2. Nevrale nettverk omtales ofte som *black box*-modeller (Chollet, 2021, s. 261; Kirkos, 2015). Det betyr at man ikke kan se nøyaktig hva modellen gjør og hvilke sammenhenger den finner. Dette fører til at tolkningen av denne typen modeller kan være intrikat. Hovedideen bak de nevralt nettverkene er imidlertid å la alle inputvariablene interagere med hverandre gjennom lange og komplekse læringskretser (Russel & Norvig, 2021, s. 801).



Figur 2: Illustrasjon av et dypt nevralt nettverk

Dyplæring er i utgangspunktet en betegnelse som brukes om nevralt nettverk med to eller flere skjulte lag. Dybden i modellen refererer til hvor mange sammenhengende skjulte lag modellen har. I senere tid har det gått inflasjon i begrepet dyplæring, og noen omtaler derfor nevralt nettverk med få lag som *shallow learning* (Chollet, 2021, s. 7). Med dagens teknologi kan det brukes flere hundre skjulte lag i nevralt nettverksmodeller. Denne oppgaven betegner et nevralt nettverk som dypt dersom det består av flere enn ett skjult lag.

Dyplæring har blitt mer og mer tatt i bruk de siste årene, mye på grunn av stadig økende mengder data og kraftigere datamaskiner som kan trene mer komplekse modeller. Dype nevrale nettverk har i prinsippet få begrensninger når det kommer til anvendelser, men de mest brukte områdene i dag er tekst-, tale- og bildegjenkjenning, selvkjørende biler og oversettelse (Chollet, 2021, s. 10-11; Russel & Norvig, 2021, s. 801). Dette er områder som det i nyere tid har blitt gjort store fremskritt innen ved bruk av dyplæring.

Nevrale nettverksmodeller kan også brukes til å predikere konkurser. Det er særlig banker og andre kreditorer som vil ha nytte av å beregne sannsynlighet for konkurs blant sine nåværende og potensielle kunder. De kan bruke slike modeller til å vurdere lånesøknader individuelt, eller se på forventet mislighold på hele utlånsporteføljen sin. I denne oppgaven vil modellene lete etter kjennetegn som mange av selskapene som går konkurs har til felles. Modellene kan deretter brukes på reelle data for å si noe om hvilke selskaper som har størst sannsynlighet for å gå konkurs i nærmeste fremtid.

1.3 Problemstilling

Det er mange som har sett på bruken av dyplæring til konkursprediksjon tidligere. Dette beskrives nærmere i kapittel 2. Den tidligere forskningen på området viser at denne typen modeller kan oppnå både like gode og bedre resultater enn det man har fått med andre metoder (Bellovary et al., 2007; Paraschiv et al., 2021; Perez, 2006). Det vises imidlertid også at man ikke har noe grunnlag for å anta at disse modellene skiller seg ut og er signifikant bedre enn andre modeller (Paraschiv et al., 2021). En mulig grunn til dette kan være de store variasjonene i hvordan dyplæringsmodeller har blitt satt opp og brukt i den tidligere forskningen. Dette utdypes nærmere i seksjon 2.4 og 2.5. Basert på dette er det interessant å se nærmere på oppsett av dyplæringsmodeller innenfor konkursprediksjon.

Nøkkeltall som beskriver et selskaps finansielle situasjon er mye brukt i den tidligere forskningen, og det er få studier som benytter regnskapspostene som egne variabler. Det vil derfor undersøkes om man kan oppnå bedre resultater ved å benytte regnskapsposter direkte i de nevrale nettverksmodellene. Problemstillingen oppgaven vil undersøke er dermed:

I hvilken grad kan ubehandlede regnskapsdata forbedre konkursprediksjon med dyplæringsmodeller?

Oppgaven vil forsøke å besvare problemstillingen ved å sette opp modeller med mye brukte nøkkeltall og sammenlikne disse med modeller som består av ubehandlede regnskapsdata. Det vil også bli laget modeller som inneholder en kombinasjon av regnskapsdata og nøkkeltall. I løpet av oppgaven vil det også bli undersøkt hvorvidt det er nødvendig å benytte finansielle nøkkeltall i konkursprediksjonsmodeller laget med nevrale nettverk.

Gode svar på problemstillingen vil kunne være til stor nytte for banker og andre kredittinstitusjoner. Mislighold av gjeld er en stor kostnad for parten som låner ut, og det vil derfor være verdifullt å forbedre modellene som ligger til grunn for vurdering av selskapers økonomiske situasjon. I praktisk bruk av konkursprediksjon er det forskjell på kostnadene for en bank ved ulike type feil i en konkursprediksjonsmodell. Det å predikere konkurs for et selskap som ikke går konkurs, vil kunne føre til at banken taper en kunde, mens det å unnlate å predikere konkurs for et selskap som går konkurs, vil kunne føre til mislighold av lån. Derfor er det viktig å ha en modell som oftere predikerer at et selskap går konkurs om det er tvil. Det finnes flere måter å forbedre en modell på, og i denne oppgaven vil hovedfokuset være på å finne variabler som gir modellen bedre prediksjonsevne. Det vil også diskuteres hvilke evalueringsmål som avgjør om en modell er god eller ikke. Denne diskusjonen vil ta utgangspunkt i de ulike kostnadene ved ulike typer statistiske feil innenfor konkursprediksjon, ettersom det viktigste for mange aktører er å oppdage flest mulig av de faktiske konkursene.

Vinklingen på denne oppgaven er inspirert av resultater og forslag fra tidligere forskning innenfor samme fagfelt. Oppgavens hovedbidrag til litteraturen om konkursprediksjon er at den gir en økt innsikt i hvordan det å benytte regnskapsdata direkte i en dyplæringsmodell kan påvirke modellens prediksjonskraft, og om dette kan erstatte eller delvis erstatte de finansielle nøkkeltallene som er blitt benyttet i denne typen modeller tidligere.

1.4 Oppgavens struktur

I kapittel 2 presenteres en oversikt over relevante elementer i den tidligere forskningen på fagfeltet. Videre blir behandlingen av datagrunnlaget gjennomgått i kapittel 3. Her presenteres også de ulike datasettene som benyttes i oppgaven. I kapittel 4 blir estimeringsteknikkene, samt evalueringskriteriene til modellene gjennomgått. Resultatene fra modellene presenteres i kapittel 5, og blir tolket og drøftet i kapittel 6. I kapittel 7 fremlegges konkluderende bemerkninger, oppgavens utfordringer og begrensninger, samt forslag til videre forskning.

2 Tidligere forskning

I dette kapitlet vil tidligere forskning innen konkursprediksjon og dyplæring gjennomgås. De første konkursprediksjonsmodellene ble tatt i bruk på 1930-tallet (Bellovary et al., 2007). På denne tiden var det univariate modeller som var mest vanlig. Disse modellene var vesentlig enklere enn dagens modeller, og krevde derfor langt mindre datakraft. Det var først i 1968 at en av de mest kjente multivariate modellene ble lansert av Edward I. Altman. Etter dette har det vært mange som har forsket videre på konkursprediksjonsmodeller, og i den nyere forskningen har bruken av dyplæring blitt mer sentralt. Videre presenteres noen av de viktigste bidragene på fagfeltet.

2.1 Altman-modellene

Z-score modellen ble utviklet av Edward I. Altman i 1968 og er regnet som en av de mest sentrale modellene innenfor konkursprediksjon (Bellovary et al., 2007). Altman lanserte denne multivariate modellen som følge av at univariate modeller ble sett på som usofistikerte og for enkle av akademikere (Altman, 1968). En multivariat modell kan se på flere nøkkeltall samtidig, og kan dermed vurdere samvariasjonen mellom disse. Dette er noe man ikke har mulighet til å gjøre med de univariate modellene som kun behandler ett nøkkeltall om gangen. Z-score modellen er en diskriminantanalyse som lager en lineærkombinasjon av variablene for å skille konkursselskaper fra ikke-konkursselskaper. Altman benyttet fem variabler bestående av finansielle nøkkeltall i denne modellen:

$$\begin{aligned} X_1 &= \frac{\text{arbeidskapital}}{\text{totalkapital}} \\ X_2 &= \frac{\text{opptjent egenkapital}}{\text{totalkapital}} \\ X_3 &= \frac{\text{resultat før rentekostnad og skatt}}{\text{totalkapital}} \\ X_4 &= \frac{\text{egenkapitalens markedsverdi}}{\text{samlet gjeld}} \\ X_5 &= \frac{\text{salgsinntekter}}{\text{totalkapital}} \end{aligned}$$

Altman et al. (1977) utviklet senere en ny modell som fikk navnet Zeta-modellen. Dette er en modell som predikerer konkurs med fokus på større selskaper. Z-score modellen og Zeta-modellen har nokså lik prediksjonsevne på selskaper som er under to år forut for konkurs, men Zeta-modellen har vesentlig bedre prediksjonsevne når det benyttes en tidshorisont på to år eller mer (Altman et al., 1977).

Altman reviderte i år 2000 den opprinnelige modellen fra 1968 (Altman, 2013), da han ønsket å lage en modell som ikke ble påvirket av størrelsen og bransjen til selskapene i like stor grad. Denne nye modellen ble kalt den reviderte fire-variabel Z-score modellen, ettersom variabel X_5 ble utelatt fra denne modellen. Det viste seg i den opprinnelige Z-score modellen at denne variabelen kunne

bli påvirket av forskjellen i selskapenes størrelse og bransje, og siden variabelen ikke hadde signifikant påvirkning på resultatene ble den fjernet i den reviderte modellen. Resultatene fra studien viser imidlertid at den reviderte fire-variabel Z-score modellen er noe mindre pålitelig enn den originale Z-score modellen (Altman, 2013).

2.2 SEBRA-modellene

SEBRA er en forkortelse for System for EDB-Basert regnskapsanalyse (Bernhardsen & Larsen, 2007). Dette er en empirisk regnskapsmodell som er utviklet av Norges Bank. Modellen ble laget for å analysere konkurssannsynlighetene for ulike aksjeselskap i Norge, samt bankenes forventede utlånstap.

Den opprinnelige SEBRA-modellen har i stor grad nøkkeltall som variabler. Variablene inneholder blant annet informasjon om inntjening, likviditet og soliditet (Eklund et al., 2001). Det er naturlig å anta at et selskaps evne til å oppfylle forpliktelser på kortere og lengre sikt kan si noe om sannsynlighet for konkurs. De øvrige variablene i modellen inneholder informasjon om alder, størrelse og bransje. Dette er tilleggsinformasjon som kan være nyttig for prediksjonen, da det er store variasjoner på tvers av bransjer og mellom store og små selskaper. Bransjevariablene beregnes ut fra gjennomsnitt og varians for bransjen i de finansielle nøkkeltallene som er brukt i modellen (Bernhardsen, 2001). Dette gjør at modellen kan ta hensyn til forskjeller i konkurrisisiko i ulike bransjer.

Bernhardsen og Larsen (2007) har utviklet to nye versjoner av SEBRA-modellen. Den første videreutviklede SEBRA-modellen kalles SEBRA-basis. En av de viktigste forskjellene i denne modellen sammenliknet med den opprinnelige modellen er at variabelen for størrelse er fjernet. I tillegg blir bransjevariablene beregnet årlig i stedet for over hele tidsperioden man modellerer for. SEBRA-basis er altså en forenkling av den opprinnelige modellen. Dette fører til at store selskaper, som i den opprinnelige modellen fikk en systematisk lavere sannsynlighet for konkurs, ikke ble vurdert til å ha en kunstig lav risiko for konkurs (Bernhardsen & Larsen, 2007). Det at mindre selskaper har større risiko for å gå konkurs var ett av momentene som ble diskutert av Bernhardsen og Larsen (2007) ved utarbeidelsen av denne nye modellen. Den andre videreutviklede SEBRA-modellen kalles SEBRA-utvidet. Denne modellen tar i tillegg til variablene i SEBRA-basis med variabler for størrelse, leverandørgjeld og skyldige offentlige avgifter. Dette er variabler som i stor grad er knyttet til selskapets størrelse (Bernhardsen & Larsen, 2007).

SEBRA-utvidet egner seg ifølge Bernhardsen og Larsen (2007) best til prediksjoner på foretaksnivå, mens SEBRA-basis egner seg best til å regne ut bankenes forventede utlånstap. Den forenklete modellen har som nevnt ikke med variablene som varierer mye ut fra selskapets størrelse. Analyser av bankers forventede utlånstap vil naturlig nok vektlegge mislighold fra større selskaper mer enn fra små, siden dette vil føre til at banken relativt sett taper mer penger. Store selskaper med dårlige finansielle nøkkeltall vil i den opprinnelige modellen i større grad bli vurdert

til lavere sannsynlighet for konkurs, fordi større selskaper har større marginer. Dette fører til at basis-modellen egner seg best til å analysere utlånstap, mens den utvidede modellen egner seg bedre til konkursprediksjon hvor størrelsesvariablene ikke systematisk påvirker alle observasjonene i modellen.

2.3 Andre metoder for konkursprediksjon

Det finnes ingen enkeltmetoder for konkursprediksjon som gjør det best på alle områder (Wang et al., 2014). Dette kan skyldes at det er store variasjoner i den tidligere forskningen med tanke på strukturen i datasettet, detaljer i problemet som skal løses og hvordan modellene settes opp og brukes.

En artikkel skrevet av Ravi Kumar og Ravi (2007) tar for seg forskning på konkursprediksjon fra 1968-2005. I artikkelen er det samlet mange ulike modeller som har blitt utviklet på dette området, og det legges særlig vekt på den økende bruken av maskinlæringsteknikker. Nevrale nettverk er en av de metodene som brukes i flest av artiklene det vises til i denne artikkelen, noe som tyder på at det er en mye brukt metode innenfor forskningen på konkursprediksjon. Dette bekreftes også av Bellovary et al. (2007) som viser til at bruken av nevralt nettverk har økt betraktelig.

Perez (2006) kommer i sin litteraturstudie også frem til at nevralt nettverk presterer veldig bra ved anvendelse på konkursprediksjonsproblemer. Artikkelen tar for seg 30 ulike studier på klassifisering av konkurser ved bruk av nevralt nettverk mellom 1990 og 2000. I flertallet av disse studiene er konklusjonen at resultatene fra de nevralt nettverkene er minst like gode som flere av de tradisjonelle metodene som har vært mye brukt til konkursprediksjon. Det trekkes blant annet frem at nevralt nettverk presterer like bra eller bedre enn multivariate diskriminantanalyser og logistiske regresjonsmodeller (Perez, 2006). En studie av Paraschiv et al. (2021) viser også at nevralt nettverk presterer bedre enn logistiske regresjoner, men at det ikke er signifikante forskjeller mellom de to metodene.

I en artikkel av Næss et al. (2017) undersøkes det videre på om maskinlæringsteknikker er i bedre stand til å estimere modeller for konkursprediksjon enn tradisjonelle metoder. Ved å bruke flere ulike metoder på et stort norsk datasett kommer de frem til at det er generalisert additiv modell, en utvidelse av en generalisert lineær modell, som egner seg best til formålet. Artikkelen viser også at nevralt nettverk er den best egnede maskinlæringsteknikken. Dette er uavhengig av hvilke av de testede variabelsettene som brukes i modellene. I artikkelen kommer det frem at generalisert additiv modell ga marginalt bedre prediksjon enn nevralt nettverk. Forskjellen mellom de to metodene er dog ikke signifikant. Det vises også at forskjellen mellom metodene blir mindre jo flere variabler som brukes i modellen, og Næss et al. (2017) foreslår derfor å forsøke å bruke flere variabler i arbeidet med nevralt nettverk til konkursprediksjon.

Siden design av nevrale nettverk og valg av uavhengige variabler er noe det i tidligere forskning er mange ulike tilnæringer til, vil de neste seksjonene ta for seg noen artikler som er relevante for å gi et godt svar på problemstillingene i denne oppgaven.

2.4 Design av nevrale nettverksmodeller i tidligere forskning

Det har lenge vært forsket på nevrale nettverk. En artikkel skrevet av Zhang et al. (1999) drøfter potensialet til nevrale nettverk som en konkurrerende teknikk til mer tradisjonelle statistiske metoder. I denne artikkelen oppfordres det til å forske videre på oppsettet av nevrale nettverksmodeller, da mer kunnskap om hvordan disse modellene fungerer potensielt vil kunne bidra til bedre resultater. Her pekes det spesielt på oppsettet av selve nettverket, med antall skjulte lag og antall noder som eksempler. Dette er parametere som det i dagens forskning fortsatt er usikkerhet rundt, og det finnes ingen standard for hvordan disse bør settes opp (Chollet, 2021, s. 145). I den tidligere forskningen på konkursprediksjon ved bruk av nevrale nettverk er det derfor store forskjeller i hvordan modellene er utformet.

Til tross for at det ikke finnes en felles struktur for design og oppsett av de nevrale nettverksmodellene, er det vanlig å ha samme antall noder i inputlaget som antallet variabler i datasettet. Dette vises i mye av den tidligere forskningen (Næss et al., 2017; Odom & Sharda, 1990; Paraschiv et al., 2021; Wilson & Sharda, 1994). Det benyttes også som regel en eller to noder i outputlaget når man jobber med konkursprediksjon (Perez, 2006). Mye av den tidligere forskningen denne oppgaven bygger på bruker én outputnode (Odom & Sharda, 1990; Paraschiv et al., 2021; Pelja & Wahlstrøm, 2021), men to noder brukes blant annet av Næss et al. (2017); Wilson og Sharda (1994). Det er større variasjon når det kommer til antall noder i de skjulte lagene, men det benyttes eksempelvis 5 (Odom & Sharda, 1990; Paraschiv et al., 2021), 10 (Næss et al., 2017; Wilson & Sharda, 1994) og 20 (Pelja & Wahlstrøm, 2021) noder i disse lagene i den tidligere forskningen.

Når det kommer til antall skulte lag er det også noen variasjoner i den tidligere forskningen. Mange modeller benytter kun ett skjult lag (Næss et al., 2017; Odom & Sharda, 1990; Wilson & Sharda, 1994), og basert på definisjonen i denne oppgaven kvalifiserer ikke disse som dyplæringsmodeller. Andre studier benytter nevrale nettverk med flere skjulte lag. Eksempelvis bruker Paraschiv et al. (2021) to skjulte lag og Pelja og Wahlstrøm (2021) bruker tre skjulte lag. Dette underbygger at det er forskjeller i hvordan modelldesignet er i den tidligere forskningen, også for modeller som baseres på tilnærmet like datasett. De fleste relevante eksemplene i den tidligere forskningen er satt opp med relativt få skjulte lag, og dette kan potensielt forklares ved at enklere modeller har mindre problemer knyttet til overtilpasning enn det mer komplekse modeller har (Chollet, 2021, s. 148).

2.5 Variabler i tidligere forskning

I den tidligere forskningen brukes det mange ulike variabler og det finnes ingen felles konsensus blant akademikerne om hvilke variabler som bør være med i konkursprediksjon (Paraschiv et al., 2021; Tian et al., 2015). Perez (2006) viser at det heller ikke er fastslått hva som gir best resultater når det kommer til antall variabler i modellen og størrelsen på datasettet som brukes. Det er i utgangspunktet opp til hver enkelt forsker å avgjøre hvilke variabler som skal benyttes i de ulike modellene. Det vises likevel at valget av variabler ikke kun er et resultat av bevisste valg fra forskernes side, da valget ofte begrenses av hvilke data forskeren har tilgang til (Perez, 2006).

Ut fra det tilgjengelige datagrunnlaget kan forskerne benytte flere ulike metoder for å plukke ut variabler til modellene. En mulighet er å gjennomføre en tilfeldig utvelgelse av variabler til modellen, men det vises blant annet i Paraschiv et al. (2021) at dette er en dårlig strategi. Mange velger å benytte variabelsett som er utarbeidet av andre anerkjente forskere, og som dermed er mye brukt innenfor fagfeltet allerede. Z-score-modellen til Altman (1968) er et eksempel på et slikt variabelsett. Selv om dette er vanlig praksis, mener Begley et al. (1996) at noen av de tradisjonelle variabelsettene er utdaterte. I artikkelen kommer det frem at disse ikke lenger foretrekkes i modeller for konkursprediksjon nå som teknologien og kunnskapen er bedre enn det den var for noen tiår siden. Selv om Ravi Kumar og Ravi (2007) viser at forskningen på fagfeltet har begynt å benytte andre variabler i arbeidet med konkursprediksjon er de fem variablene til Altman fortsatt mye brukt, både alene og sammen med andre variabler (Kainth & Wahlstrøm, 2021; Næss et al., 2017; Pelja & Wahlstrøm, 2021; Zhang et al., 1999). Også SEBRA-variabelsettene til Bernhardsen og Larsen (2007) er velkjent i den norske litteraturen, og har blitt brukt i mye tidligere forskning (Kainth & Wahlstrøm, 2021; Næss et al., 2017; Pelja & Wahlstrøm, 2021).

Forskere har også mulighet til å bruke algoritmer for å velge ut de beste variablene til konkursprediksjonsmodellene. Denne formen for variabelutvelgelse er mye brukt innen maskinlæring, og det finnes mange metoder med ulike algoritmer å velge mellom. *LASSO* og *Ridge regression* er mye brukte eksempler på slike metoder. Det vises i den tidligere forskningen at denne måten å velge ut variabler på kan oppnå bedre resultater enn ved å benytte datasett som er utarbeidet av andre (Paraschiv et al., 2021; Tian & Yu, 2017). Perez (2006) peker på mulighetene ved å benytte nevrale nettverksmodeller, og mener at det kan være hensiktsmessig å la denne typen modeller velge ut de beste variablene selv.

Et annet moment ved valg av variabler til konkursprediksjon er at det å inkludere mange uavhengige variabler kan øke kompleksiteten til modellen (Paraschiv et al., 2021). Dette vil kunne føre til at modellen mister ytelse, ettersom kompleksitet og ytelse er en avveining i slike modeller (Paraschiv et al., 2021). En mer kompleks modell vil kunne gi mer informasjon om det den skal predikere, men hvis ytelsen blir for dårlig er det lite hensiktsmessig å bruke modellen i praksis.

2.6 Kategorier av nøkkeltall

De ulike finansielle nøkkeltallene kan samles i kategorier som sier noe om ulike forhold knyttet til et selskaps økonomiske situasjon. Her presenteres noen av de mest sentrale kategoriene innenfor konkursprediksjon.

Lønnsomhet

Lønnsomhet er et begrep som brukes om et selskaps inntjening. Inntekter som er høyere enn kostnadene skaper et overskudd og er et tegn på god lønnsomhet. Lønnsomhet målt ved bruk av regnskapsdata kalles rentabilitet (Langli, 2016). Vanlige mål er egenkapitalrentabilitet og total kapitalrentabilitet. Disse målene er et uttrykk for utviklingen i henholdsvis egenkapitalen og total kapitalen fra en periode til en annen. God lønnsomhet er viktig for at et selskap skal kunne være levedyktig på lengre sikt.

Likviditet

Likviditet er et mål på et selskaps likvide midler. Likviditet er et mål på betalingsevne da det sier noe om evnen til å betale løpende forpliktelser (Langli, 2016). Det er store forskjeller mellom ulike bransjer på hvor mye likvide midler et selskap behøver å ha, men felles for de aller fleste er at for dårlig likviditet fører til betalingsproblemer som igjen øker sjansen for konkurs. Derfor er dette et viktig mål innenfor konkursprediksjon.

Soliditet

Soliditet er et mål på hvor mye av et selskap som er finansiert med egenkapital. I motsetning til likviditet som ser på kortsiktig betalingsevne, handler soliditet om evnen til å håndtere forpliktelser på lengre sikt (Langli, 2016). Soliditet viser derfor et selskaps evne til å tåle perioder med dårlige resultater. Dersom et selskap har dårlig soliditet, er det finansiert med mye gjeld.

Alder

Alderen til et selskap er relevant for konkursprediksjon, siden mange selskaper går konkurs i løpet av sine første leveår (Statistisk sentralbyrå, 2021). Aksjeselskap er den mest overlevelsesdyktige organisasjonsformen, og aksjeselskapene som ble etablert i Norge i 2013 har hatt en overlevelsesprosent på 46 % i løpet av de første fem årene etter etablering (Statistisk sentralbyrå, 2020). Selskap som har eksistert lenge har gjerne, i motsetning til yngre selskap, større grad av stabilitet i driften med en solid markedsandel og kundegruppe. Statistikken tyder på at selskapene ofte går konkurs før de rekker å bli godt etablert i markedet.

3 Data

I dette kapitlet presenteres datagrunnlaget og behandlingen av dette. De ulike variabelsettene som benyttes i oppgaven blir gjennomgått, etterfulgt av hvordan disse har blitt behandlet før kjøring av modellene. Kapitlet avsluttes med en diskusjon rundt kvaliteten på datagrunnlaget.

3.1 Beskrivelse og behandling av datagrunnlag

Datagrunnlaget i denne oppgaven er utarbeidet av Wahlstrøm (2022) og har blitt brukt i mye av hans tidligere forskning (Kainth & Wahlstrøm, 2021; Paraschiv et al., 2021; Pelja & Wahlstrøm, 2021). Data er hentet fra Brønnøysundregisteret i 2021 og består av totalt 3 876 923 ikke-konsoliderte norske årsregnskaper fra regnskapsårene 2006 til 2019. Det ubehandlede datagrunnlaget består av 286 variabler som viser til ulike poster i årsregnskapet og annen relevant informasjon knyttet til selskapene.

Det er hensiktsmessig å gjennomføre mye av den samme databehandlingen og de samme avgrensningene som har blitt gjort i tidligere relevant forskning. Dette er viktig for å kunne sammenlikne modellene i denne oppgaven med modeller i tidligere forskning på et så godt grunnlag som mulig. Dette er særlig relevant da oppgavens formål er å komme med en forbedring av de eksisterende prediksjonsmodellene.

Datagrunnlaget har blitt filtrert til å kun inneholde alle norske private, ikke-børsnoterte aksjeselskap. Informasjonen til disse selskapene er lett tilgjengelig, da disse selskapene er pålagt å rapportere sine regnskap til myndighetene. Dette står i motsetning til blant annet enkeltpersonforetak som i mye større grad har frivillig rapportering. Datagrunnlaget består av kun selskapsregnskaper, ettersom regnskap fra morselskapet i et konsern er ekskludert for å unngå at noen regnskapstall blir analysert flere ganger. Dette gjøres i henhold til Næss et al. (2017).

Datagrunnlaget har også flere variabler hvor alle verdier er null, da disse tilsvarer overskriftsråder i regnskapet. Det er ingen hensikt i å benytte disse til konkursprediksjon og de variablene dette gjelder for har derfor blitt fjernet. Flere av variablene i datagrunnlaget har ikke registrerte verdier for alle årsregnskaper. For de variablene hvor manglende registreringer overstiger 95 % av alle observasjonene fjernes variablene fra datagrunnlaget. Dette gjøres fordi variabler med under 5 % faktiske verdier kan gi uønskede resultater. Et eksempel på dette er at samtlige verdier i en variabel kan blir satt til null ved gjennomføring av *winsorizing* slik det gjøres i denne oppgaven.

Standard for næringsgruppering er en mye brukt statistisk standard som bygger på EUs næringsstandard NACE (Statistisk sentralbyrå, u.å.). Observasjoner med NACE-koder som begynner på 64, 65 og 66 ekskluderes fra datagrunnlaget. Dette tilsvarer observasjoner innenfor næringshovedområde K som omfatter næringer innen *finansierings- og forsikringsvirksomhet* (Statistisk sentralbyrå, u.å.). På samme måte som i Næss et al. (2017), Mansi et al. (2012) og

Bernhardsen og Larsen (2007), ekskluderes disse for å kun ha ikke-finansielle aksjeselskaper i datagrunnlaget. Mansi et al. (2012) fjerner også observasjoner innenfor næringshovedområdene D, E og L, som viser til henholdsvis næringene *elektrisitets-, gass-, damp- og varmtvannsforsyning, vannforsyning, avløps- og renovasjonsvirksomhet* og *omsetning og drift av fast eiendom* (Statistisk sentralbyrå, u.å.). Dette har blitt en vanlig ekskludering i den tidligere forskningen på konkursprediksjon (Kainth & Wahlstrøm, 2021; Pelja & Wahlstrøm, 2021), og disse næringene fjernes derfor også i denne oppgaven. I likhet med Næss et al. (2017), fjernes observasjoner med NACE-koder som begynner på 41 for å sikre relevant sammenlikningsgrunnlag. Observasjoner med manglende NACE-kode har også blitt fjernet fra datagrunnlaget.

Årsregnskaper hvor det er mer enn 5 % avvik i balansen har blitt fjernet. Disse fjernes da det skal være balanse mellom egenkapital og gjeld, og summen av totale eiendeler i regnskapet. Når det er avvik mellom disse postene tyder det på det har skjedd noe feil, eksempelvis i innlasting og konvertering av datagrunnlaget. I likhet med tidligere forskning ekskluderes årsregnskaper hvor verdien av regnskapsposten *sum eiendeler* er under 500 000 kroner fra datagrunnlaget (Bernhardsen & Larsen, 2007; Næss et al., 2017; Pelja & Wahlstrøm, 2021).

Det er relevant å undersøke hvor raskt en konkurs inntreffer etter siste registrerte årsregnskap. Fra datagrunnlaget som benyttes i denne oppgaven vises det at for selskap som er klassifisert som konkurs så går 24,47 % konkurs innen ett år, 86,37 % innen to år og 98,89 % innen tre år. Siden det ofte går tre år før registrering av konkurs etter siste avlagte årsregnskap (Bernhardsen, 2001) fjernes alle årsregnskapene fra regnskapsårene 2018 og 2019. Dette gjøres da 2020 er det siste hele året med registrerte konkursdatoer. Dette støttes av mye tidligere litteratur (Bernhardsen & Larsen, 2007; Næss et al., 2017; Pelja & Wahlstrøm, 2021), og gjøres for at selskap som er i ferd med å gå konkurs ikke feilaktig skal klassifiseres som ikke-konkurs. Den siste registrerte konkursdatoen i datagrunnlaget er 2.6.2021. Konkursdatoene hentes derfor frem til og med denne datoen, noe som betyr at selskap som går konkurs i 2017 har over 98,89 % sannsynlighet for å bli klassifisert som konkurs i vårt datagrunnlag.

Etter databehandlingen består datagrunnlaget av totalt 1 067 243 årsregnskaper fra 182 206 selskaper fordelt på tolv år.

3.2 Avhengig variabel

Den avhengige variabelen i denne oppgaven er en konkursvariabel som angir hvorvidt et selskap har gått konkurs eller ikke. Konkursvariabelen er binær og kan kun ha verdiene 0 og 1. Verdien 0 brukes dersom et selskap er klassifisert som ikke-konkurs og verdien 1 brukes for klassifisering som konkurs.

Den avhengige variabelen har noen forutsetninger som må oppfylles for at et selskap skal klassifiseres som konkurs. I denne oppgaven benyttes den samme definisjonen som sentralbanken i Norge benytter i arbeidet med sin SEBRA-modell (Bernhardsen & Larsen, 2007). Det første kriteriet er at selskapet må ha sluttet å levere årsregnskap. Dette betyr at det kun er det siste registrerte årsregnskapet til et selskap som kan klassifiseres som konkurs. Det andre kriteriet er at selskapet må ha blitt erklært konkurs, enten av seg selv eller av kreditorer gjennom en domstol. Dersom begge disse kriteriene er oppfylt samtidig, får selskapets siste leverte årsregnskap status som konkurs i datagrunnlaget som brukes i denne oppgaven, og den avhengige variabelen får verdien 1. Den avhengige variabelen til de resterende årsregnskapene får verdien 0.

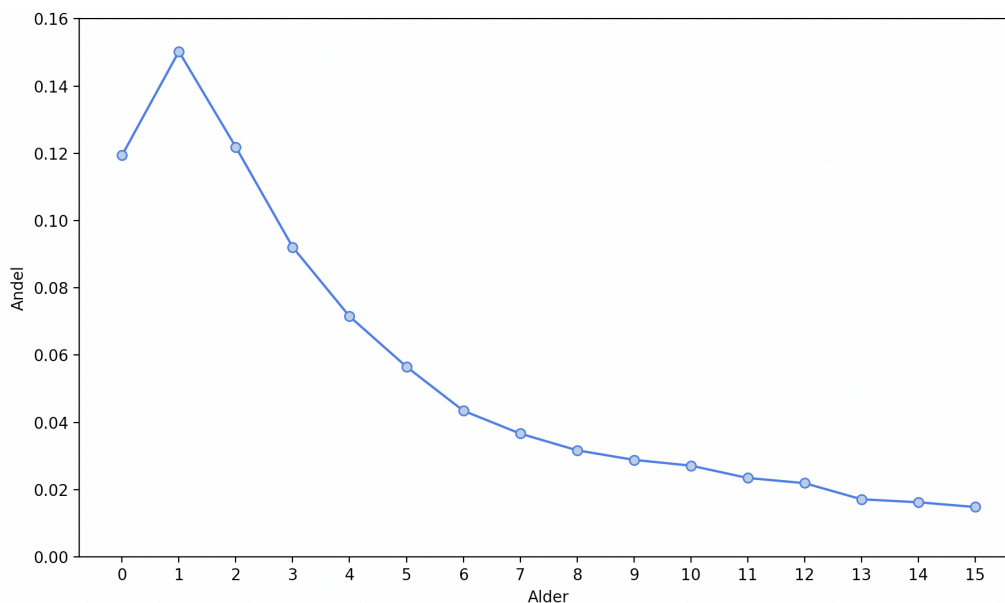
Ved bruk av denne definisjonen av konkurs består datagrunnlaget av 16 895 konkurser og 1 050 348 ikke-konkurser. Dette betyr at 1,58 % av de observerte årsregnskapene er klassifisert som konkurs. Fordelingen av konkurser per år i datagrunnlaget vises i tabell 1. Selskaper som har gått konkurs i det oppgitte året vises som konkurs, mens selskaper som går konkurs ved et senere tidspunkt vil i de foregående årene betegnes som ikke-konkurs.

Tabell 1: Fordeling av konkurser per år i datagrunnlaget i perioden 2006-2017

År	Antall selskap	Konkurs	Ikke-konkurs	Andel konkurs
2006	81 013	1 194	79 819	1,47 %
2007	82 359	1 807	80 552	2,19 %
2008	82 731	1 631	81 100	1,98 %
2009	82 350	1 483	80 867	1,80 %
2010	83 022	1 270	81 752	1,53 %
2011	84 667	1 330	83 337	1,57 %
2012	87 928	1 394	86 534	1,59 %
2013	90 815	1 366	89 449	1,50 %
2014	93 741	1 348	92 393	1,44 %
2015	96 595	1 229	95 366	1,23 %
2016	99 736	1 449	98 287	1,45 %
2017	102 286	1 394	100 892	1,36 %
Totalt	1 067 243	16 895	1 050 348	1,58 %

Tidstrenden i datagrunnlaget blir borte ved at årsregnskapene til et selskap klassifiseres som ikke-konkurs i årene før en konkurs inntreffer. Dette fører til at selskapenes utvikling over flere år ikke blir tatt hensyn til i modellene. Denne måten å klassifisere konkurser på er velkjent i litteraturen innen økonomi, og metoden brukes også i den tidligere forskningen på konkursprediksjon med dyplæring (Næss et al., 2017; Pelja & Wahlstrøm, 2021). Denne klassifiseringen benyttes derfor også i denne oppgaven, til tross for at Kirkos (2015) mener det er mer logisk å behandle konkurser som langsgående prosesser, ettersom et selskaps finansielle situasjon gradvis endres over tid.

I figur 3 kan man se at det er en overvekt av konkurser for de yngste selskapene. Dette stemmer som nevnt i seksjon 2.6 med observasjonene fra Statistisk sentralbyrå (2021), og tyder på at datagrunnlaget i oppgaven representerer den ekte fordelingen av konkurser basert på hvor lenge et selskap har eksistert.



Figur 3: Andel av totale konkurser i datagrunnlaget basert på selskapsalder

3.3 Uavhengige variabler

Som vist i seksjon 2.5 er det ingen samstemmighet rundt hvilke variabler som skal benyttes til konkursprediksjon, og det benyttes derfor mange ulike variabelsett i tidligere studier. På tross av at det ikke finnes noen veiledende teori om hvilke variabler som gir best resultater, vises det likevel at en del av de samme variablene går igjen i mye av forskningen innenfor fagfeltet. Mange av de variablene som er mye brukt er nøkkeltall. Disse er ofte hentet fra økonomisk teori og sier noe om den generelle økonomiske situasjonen til et selskap. Flere av disse variablene kan samles i felles kategorier av nøkkeltall som beskrevet i seksjon 2.6.

I denne oppgaven blir det kun benyttet kvantitative inputvariabler, noe som også er tilfellet for mye sammenlignbar forskning. Ifølge Perez (2006) er dette vanlig i denne typen forskning da denne informasjonen er lett tilgjengelig. De uavhengige variablene i denne oppgaven er hentet fra regnskapsdata som beskrevet i seksjon 3.1. Variablene består i hovedsak av regnskapspostene benyttet direkte i modellen, men også ulike finansielle nøkkeltall. Flertallet av nøkkeltallene som anvendes er forholdstall, noe som vektlegger forholdet mellom variablene mer enn selve tallverdien. Dette er med på å normalisere dataene slik at de blir sammenlignbare på tvers av ulike tidspunkter, samt at det blir enklere å sammenlikne selskaper med ulik størrelse.

Det benyttes seks ulike variabelsett i oppsettet av modellene i oppgaven. De valgte variabelsettene beskrives i de påfølgende seksjonene, og en oversikt over variablene i de ulike variabelsettene finnes i appendiks A1.

3.3.1 Variabelsett **Altman**

Det første variabelsettet som benyttes i denne oppgaven er hentet fra Altmans Z-score-modell (Altman, 1968). Tabell 2 viser nøkkeltallene som utgjør variabelsett **Altman**. Disse fem uavhengige variablene er bygd opp av poster i regnskapet, og utarbeidelsen av dem vises i appendiks A2. Som vist i seksjon 2.5 er dette variabelsettet mye benyttet i den tidlige forskningen (Kainth & Wahlstrøm, 2021; Næss et al., 2017; Pelja & Wahlstrøm, 2021; Zhang et al., 1999), og det benyttes i denne oppgaven for å lettere kunne vurdere resultater på tvers av de ulike studiene. I den originale Z-score-modellen benyttet Altman (1968) markedsverdien av egenkapitalen, men denne erstattet han senere med bokført verdi for å tilpasse modellen til ikke-børsnoterte selskaper (Altman, 1993). Ettersom denne oppgaven baseres på data fra ikke-børsnoterte selskaper benyttes det her bokført verdi av egenkapital. Dette er også i henhold til det som gjøres i Pelja og Wahlstrøm (2021).

*Tabell 2: Variabler i variabelsett **Altman***

Variabelnavn	Beskrivelse	Type
X ₁	$\frac{\text{arbeidskapital}}{\text{eiendeler}}$	Kontinuerlig
X ₂	$\frac{\text{opptjent egenkapital}}{\text{eiendeler}}$	Kontinuerlig
X ₃	$\frac{\text{EBIT}}{\text{eiendeler}}$	Kontinuerlig
X ₄	$\frac{\text{egenkapital}}{\text{gjeld}}$	Kontinuerlig
X ₅	$\frac{\text{salgsinntekter}}{\text{eiendeler}}$	Kontinuerlig

3.3.2 Variabelsett **SEBRA**

Det andre variabelsettet som benyttes i oppgaven er basert på SEBRA-utvidet, som er utviklet av Bernhardsen og Larsen (2007). Dette variabelsettet velges fremfor den originale SEBRA-modellen og SEBRA-basis, da Bernhardsen og Larsen (2007) har konkludert med at SEBRA-utvidet gir bedre prediksjoner på foretaksnivå, mens SEBRA-basis gir bedre prediksjon for å beregne bankers aggregerte utlånstap. Siden det i denne oppgaven er større fokus på konkursprediksjon enn på aggregerte utlånstap, virker det mest fornuftig å velge SEBRA-utvidet.

Variabelsett **SEBRA** består av de 15 variablene som vises i tabell 3. Variablene dekker selskapets resultat, egenkapitalprosent, likviditetsprosent, eiendeler, leverandørgjeldprosent, avgiftsprosent, i tillegg til at det er åtte mulige dummyvariabler for selskapets alder i år. Dummyvariablene for alder i år er laget ved å dele variabelen *age_in_days* på 365, og deretter runde svaret ned til nærmeste hele år. Alle selskap med alder over 8 år er samlet i dummyvariabelen *age_year_8*. I tillegg inneholder variabelsettet dummyvariabelen *ek_d* som viser om et selskaps innskutte egenkapital er mindre enn den bokførte egenkapitalen. I likhet med det som gjøres av Bernhardsen og Larsen (2007) i den opprinnelige SEBRA-utvidet modellen blir det tatt logaritmen av eiendelene i variabelen *eie*.

Tabell 3: Variabler i variabelsett SEBRA

Variabelnavn	Beskrivelse	Type
<i>res</i>	Ordinært resultat før av- og nedskrivninger i prosent av total gjeld	Kontinuerlig
<i>ek</i>	Egenkapital i prosent av total kapital	Kontinuerlig
<i>ek_d</i>	1 = Innskutt egenkapital mindre enn bokført egenkapital 0 = Innskutt egenkapital større enn eller lik bokført egenkapital	Dummy
<i>likv</i>	Likvider minus kortsiktig gjeld i prosent av omsetning	Kontinuerlig
<i>eie</i>	Log(Sum eiendeler i faste kroner)	Kontinuerlig
<i>lev</i>	Leverandørgjeld i prosent av total kapital	Kontinuerlig
<i>avg</i>	Skyldige offentlige avgifter i prosent av total kapital	Kontinuerlig
<i>age_year_1</i>	1 = selskapets alder er over 1 år 0 = selskapets alder er ikke over 1 år	Dummy
<i>age_year_2</i>	1 = selskapets alder er over 2 år 0 = selskapets alder er ikke over 2 år	Dummy
<i>age_year_3</i>	1 = selskapets alder er over 3 år 0 = selskapets alder er ikke over 3 år	Dummy
<i>age_year_4</i>	1 = selskapets alder er over 4 år 0 = selskapets alder er ikke over 4 år	Dummy
<i>age_year_5</i>	1 = selskapets alder er over 5 år 0 = selskapets alder er ikke over 5 år	Dummy
<i>age_year_6</i>	1 = selskapets alder er over 6 år 0 = selskapets alder er ikke over 6 år	Dummy
<i>age_year_7</i>	1 = selskapets alder er over 7 år 0 = selskapets alder er ikke over 7 år	Dummy
<i>age_year_8</i>	1 = selskapets alder er over 8 år 0 = selskapets alder er ikke over 8 år	Dummy

3.3.3 Variabelsett **Regnskap**

De fire siste variabelsettene i oppgaven er i hovedsak basert på ubehandlede regnskapsdata fra årsregnskapene. Det første variabelsettet med regnskapsposter, **Regnskap**, er utarbeidet ved å hente ut alle regnskapspostene fra datagrunnlaget. Dette gjøres da det er ønskelig å benytte disse direkte i modellen, uten å lage nye variabler og nøkkeltall, for å se om det nevralt nettverket klarer å finne disse sammenhengene på egenhånd. Ved å bruke regnskapspostene på denne måten vil man få datasett med flere variabler, noe som ifølge Næss et al. (2017) kan være gunstig i arbeidet med nevralt nettverk. Dette variabelsettet gir modellen tilgang til mer data enn de variabelsettene som har færre variabler. Det er derfor viktig å være oppmerksom på at dette kan føre til en mer kompleks modell, noe som igjen kan gi uønsket utslag på modellens tilpasning slik som beskrevet i seksjon 2.5.

3.3.4 Variabelsett **Regnskap-**

I variabelsettet **Regnskap-** er alle summeringspostene fra **Regnskap** fjernet. Postene som fjernes i dette variabelsettet er samleposter for flere av de andre postene i regnskapet, og alle postene som har «sum» i navnet er dermed fjernet. Et eksempel på en slik post er sum innskutt egenkapital, som består av underpostene selskapskapital, overkursfond, annen innskutt egenkapital og egne aksjer i regnskapet. Resultatposter, som for eksempel ordinært resultat før skatt og driftsresultat, er ikke tatt ut fra dette variabelsettet. Disse postene ansees å være viktige, ettersom modellene ikke kan skille på hvilke poster i regnskapet som legges til eller trekkes fra i resultatpostene. I summeringspostene adderes alle de underliggende regnskapspostene, og disse kan derfor sees på som rene oppsamlingsposter. I denne oppgaven er det ønskelig å teste om de postene som kun er samleposter gir noe informasjon, og det er derfor bare summeringspostene som fjernes.

Siden summeringspostene er kombinasjoner av flere andre regnskapsposter, fjernes de i **Regnskap-** for å se om modellen klarer å finne gode sammenhenger i datasettet uten disse samlepostene. På denne måten blir modellen i større grad tvunget til å bruke de underliggende postene, og man vil kunne tolke og analysere modellen på et mer detaljert plan.

3.3.5 Variabelsett **Regnskap+Altman**

Variabelsettet **Regnskap+Altman** er likt **Regnskap** bortsett fra at variablene fra variabelsett **Altman** også er inkludert. Formålet med dette variabelsettet er å se på hvordan regnskapsposter fungerer sammen med nøkkeltall som er basert på de samme postene. Altman sine variabler er laget utelukkende basert på regnskapstall, noe som gjør at variabelsett **Altman** vil egne seg bedre til dette formålet enn for eksempel variabelsett **SEBRA**. Det er også ønskelig at dette variabelsettet skal kunne brukes til å undersøke hvor stor forklaringsstyrke Altman-nøkkeltallene har.

3.3.6 Variabelsett **Regnskap+Altman-**

Det siste variabelsettet basert på regnskapsposter, **Regnskap+Altman-**, likner i stor grad på **Regnskap+Altman**. Forskjellen mellom de to variabelsettene er at de underliggende regnskapspostene som inngår i Altman-nøkkeltallene er fjernet i **Regnskap+Altman-**. Som vist i appendiks A1, er det de åtte postene sum omløpsmidler, sum kortsiktig gjeld, sum eiendeler, sum opptjent egenkapital, ordinært resultat før skattekostnad, sum egenkapital, sum gjeld og sum inntekter som er fjernet. Disse regnskapspostene fjernes i dette variabelsettet for å kunne se på den isolerte effekten av Altman-nøkkeltallene sammen med de resterende regnskapspostene.

3.4 Utarbeidelse av datasett

De seks variabelsettene beskrevet i seksjon 3.3.1-3.3.6 sammen med konkursvariabelen i seksjon 3.2 utgjør datasettene som benyttes i modellene. I denne seksjonen gjennomgås utformingen av, samt videre behandling av datasettene. Det blir også sett på hvordan data er oppdelt i trenings-, validerings- og testsett.

3.4.1 Dele på 0

Det at en andel av verdiene som benyttes i nevneren i forholdstallene har verdien null skaper problemer i utarbeidelsen av de to første og de to siste variabelsettene. Håndteringen av dette følger en liknende fremgangsmåte som i Paraschiv et al. (2021). For observasjoner med nullverdier i både nevner og teller settes verdien av variabelen til null. Dersom nevneren er null og telleren er positiv eller negativ blir verdien på variabelen satt til henholdsvis \pm uendelig. Disse verdiene settes igjen til den største eller minste reelle verdien av forholdstallet. Deretter gjennomføres winsorizing som beskrevet i seksjon 3.4.2.

3.4.2 Winsorizing

Basert på databehandling i tidligere forskning har det blitt gjennomført en *winsorizing* av alle variablene, utenom dummyvariablene, i de ulike datasettene (Kainth & Wahlstrøm, 2021; Nyitrai & Virág, 2019; Paraschiv et al., 2021; Shumway, 2001). Dette gjøres for å redusere effekten av *outliers* i de ulike variablene i datasettene. *Winsorizing* gjøres på ulike prosentiler i den tidligere forskningen, men det varierer som regel mellom å gjennomføre *winsorizing* på 1 % nivå (Paraschiv et al., 2021; Shumway, 2001) eller 5 % nivå (Kainth & Wahlstrøm, 2021). I denne oppgaven blir *winsorizingen*, i likhet med i Kainth og Wahlstrøm (2021), gjennomført på 5 % nivå ved å sette verdien av *outliers* i variablene til den 5. og 95. prosentilen til variabelen. Dette gjennomføres ved at alle verdier under den femte prosentilen settes til verdien av den femte prosentilen, og alle verdier over den 95. prosentilen erstattes med verdien på den 95. prosentilen. Grunnlaget for dette valget er at dette viser seg å ha god effekt på liknende studier på samme datasett (Kainth &

Wahlstrøm, 2021), og at Nyitrai og Virág (2019) viser at konkursprediksjonen blir bedre ved *winsorizing* med høyere prosentiler.

3.4.3 Trenings-, validerings- og testsett

Før estimering av modellene blir de seks datasettene inndelt i trenings-, validerings- og testsett. Treningssettet er de dataene modellen trenes med, mens valideringssett brukes under treningen av modellen for å tilpasse modellen på en bedre måte. Valideringssettet tester prestasjonen til modellen på data som ikke inngår i treningssettet. Modellen tester prestasjonen på valideringssettet i hver epoke for å forbedre prediksjonsevnen. Testsett brukes for å teste modellens reelle prediksjonsevne. Observasjonene som inngår i testsettet har ikke vært med i treningen av modellen. Dette er veldig viktig, da man i utforming av modeller må være veldig nøye så man ikke tester modellen på data som inngår i treningssettet (Chollet, 2021, s. 137). Ved å benytte et eget testsett som er *out-of-sample* kan man måle prediksjonsevnen til modellen i praksis.

Dersom man estimerer modeller med små treningssett vil nevralt nettverk kunne klare å finne en perfekt tilpasning. Dette skjer fordi modellene memorerer treningssettet fremfor å lære og finne sammenhenger (Kirkos, 2015). Hvis modellen er overtilpasset vil den prestere godt under treningen, men dårlig på testsettet. Avanserte modeller som nevralt nettverk gir derfor generelt bedre resultater når de trenes ved hjelp av større datasett (Markham & Rakes, 1998). Større datasett betyr at man kan ha større treningssett, og dermed kan modellen trenes med større mengder data.

Fordelingen mellom trenings-, validerings- og testsett på konkursprediksjonsmodeller har blitt vurdert av Zhao et al. (2015). I denne artikkelen kommer det frem at den fordelingen som gir best prediksjon er å ha 80 % av det totale datasettet i treningssettet, 10 % i valideringssettet og 10 % i testsettet. Denne fordelingen benyttes derfor også for datasettene i denne oppgaven. Dette medfører at datasettene har 27032, 3379 og 3379 observasjoner i henholdsvis trenings-, validerings- og testsett.

Inndelingen i trenings-, validerings- og testsett er gjort ved tilfeldig utvelgelse til hvert datasett. Dette anbefales av Chollet (2021, s. 137) for å sikre at datasettene er representative. Ved tilfeldig utvelgelse av observasjoner i de tre ulike datasettene vil man unngå at alle årsregnskaper havner i samme datasett. Ved å dele inn på denne måten risikerer man å få årsregnskaper fra samme selskap i de ulike datasettene. Siden modellene i denne oppgaven lager prediksjoner ved å se på hvert enkelt årsregnskap som en selvstendig observasjon, antas det at dette ikke vil ha betydelig innvirkning på resultatene. Denne måten å dele inn data på gjør også at det er tilfeldig hvilket årstall observasjonene i datasettene er fra, og man kan få årsregnskaper i testsettet som er eldre enn de årsregnskapene som modellen er trent på. Det å ikke ha *out-of-time* validering og testing kan skape problemer i modeller som ser på utvikling over tid (Chollet, 2021, s. 137), men som beskrevet i seksjon 3.2 ser denne oppgaven bort fra tidstrenden, og denne måten å dele opp data på kan derfor benyttes.

3.4.4 Standardisering

I alle datasettene har det blitt gjennomført en standardisering av de uavhengige variablene (Zhao & Zhang, 2014). Alle variablene, utenom dummyvariablene, er transformert ved bruk av følgende formel:

$$x_i^* = \frac{x_i - \bar{x}}{s_x}$$

Hvor x_i^* er den standardiserte variabelen og x_i er den opprinnelige variabelen, mens \bar{x} og s_x er henholdsvis gjennomsnittet og standardavviket til alle de observerte verdiene av x_i i trenings- og valideringssettet. Den standardiserte variabelen sier noe om hvor langt den opprinnelige variabelen er fra gjennomsnittet til alle de observerte verdiene av den opprinnelige variabelen. Dersom den opprinnelige variabelen er større enn gjennomsnittet får den standardiserte variabelen en positiv verdi, og dersom den er mindre blir verdien på den standardiserte variabelen negativ. Fordelen med å standardisere variablene er at man får alle variablene over på samme måleskala, som er uavhengig av den opprinnelige måleskalaen til de ulike variablene. På denne måten forhindrer man at størrelsesforskjeller mellom ulike variabler spiller inn.

3.5 Datakvalitet

Som nevnt i seksjon 2.5 er datagrunnlaget som benyttes ofte basert på de dataene forskerne har tilgang til og ikke kun på bevisst valg fra forskernes side (Perez, 2006). Dette er også tilfellet for datagrunnlaget i denne oppgaven. Dette kan begrense mulighetene for forskerne, ved at de kan bli tvunget til å ta beslutninger basert på begrensningene i datagrunnlaget heller enn egne ønsker. Dette er imidlertid lite problematisk i denne oppgaven, da det tilgjengelige datagrunnlaget er såpass stort og variert at valgfriheten til å plukke ut den informasjonen som anses å være mest relevant er svært lite begrenset. Regnskapsdataene som benyttes ansees også å være pålitelige og objektive, ettersom det er strenge regler knyttet til regnskapsføringen og at datagrunnlaget er direkte innhentet fra Brønnøysundregisteret.

Det kan være stor variasjon mellom størrelsene på selskapene i datagrunnlaget, og Pelja og Wahlstrøm (2021) viser at en modells evne til å predikere konkurs varierer med selskapets størrelse. Ved å ikke tilpasse modellene til selskapets størrelse vil modellene kunne miste prediksjonskraft. I datagrunnlaget er størrelse tatt i betraktning ved at det har blitt gjennomført en *winsorizing* hvor ekstremverdiene i variablene blir regulert, men det kan fortsatt være store størrelsesforskjeller. Det er også gjennomført standardisering for å bedre størrelsesforholdet mellom variablene, men dette forbedrer ikke størrelsesforhold innad i en enkelt variabel.

Mange variabler kan føre til problemer med multikollinearitet i modellene. Dette er særlig relevant for datasettene hvor det kan være høy korrelasjon mellom variabler. I følge Berg (2007) trenger

ikke dette å ansees som et problem i oppgaven, ettersom oppgaven i hovedsak er ment til prediksjon. Det er likevel viktig å være klar over at multikollineariteten mellom variablene kan skape problemer ved tolkning av enkeltvariabler (Berg, 2007). Resultatene fra oppgavens SHAP-analyser bør derfor tolkes med forsiktighet ved praktisk anvendelse. I likhet med Bernhardsen (2001) antas det uavhengighet mellom observasjonene i datagrunnlaget, og individuell heterogenitet antas å ikke være et problem. Det antas et det heller ikke vil være problemer knyttet til heteroskedastisitet i modellene ettersom datautvalget er relativt stort (Perez, 2006).

4 Metode

I dette kapitlet gjennomgås den metodiske fremgangsmåten som er brukt i denne oppgaven. Det legges vekt på både bakgrunnen for metodene og oppsettet av modellene. Først forklares balanseringen av datasettet, deretter estimeringen og evalueringen av modellene. Til slutt presenteres grep som er tatt for å sikre oppgavens repliserbarhet.

4.1 Balansering av datasettet

I arbeid med konkursprediksjon er forholdet mellom antall selskaper som har gått konkurs og antall selskaper som ikke har gått konkurs ofte veldig skjevfordelt. Dette skyldes at konkurs er en relativt sjelden hendelse for et selskap. Som vist i tabell 1 er skjevfordelingen mellom de to klassene markant, og av alle de 1 067 243 årsregnskapene i datagrunnlaget er det kun 1,58 % som etterfølges av en konkurs. Denne inndelingen i konkurs og ikke-konkurs vil kunne ha stor betydning for trening og testing av konkursprediksjonsmodeller, da modellene må fungere godt på både treningsdata og testdata. Dette er viktig for at de skal kunne generaliseres slik at de fungerer godt når de benyttes på ny rådata. Det er derfor sentralt hvilket forhold det er mellom konkurs og ikke-konkurs i de virkelige dataene som modellen skal brukes på.

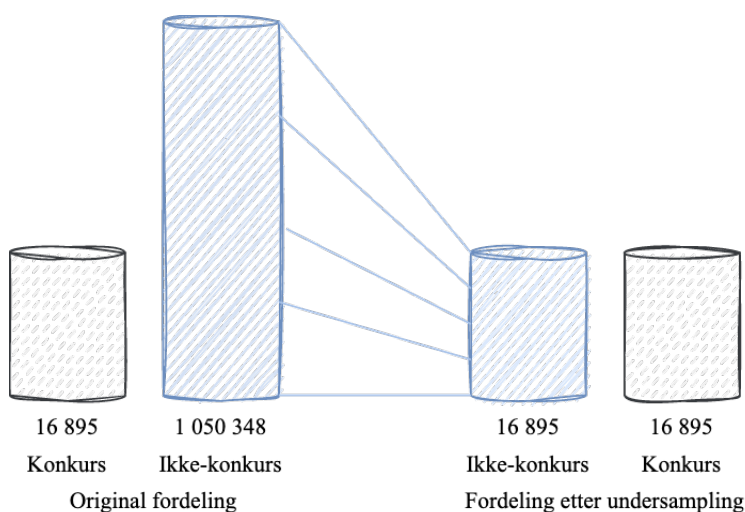
En del av den tidligere forskningen på konkursprediksjon tar for seg klassebalanseproblemet mellom de to klassene konkurs og ikke-konkurs. En måte å håndtere dette problemet på er å balansere data slik at man får en ny vektning mellom klassene. Denne håndteringen er imidlertid noe man har gått bort fra i forskningslitteraturen på konkursprediksjon. Shumway (2001) argumenterte overbevisende for å ikke balansere, og dette har etter hvert blitt standard. Det er derfor mange som velger å benytte den faktiske fordelingen av konkurs og ikke-konkurs i datasettet. Dette gjelder blant annet for Berg (2007) som har rundt 1 % konkurser i sine modeller. Det at man har gått bort fra balansering betyr imidlertid ikke at det ikke er mulig å balansere datasettet likevel, og det er fortsatt mange som gjør det.

Odom og Sharda (1990) har forsket på forholdet mellom konkurs og ikke-konkurs i datasett som brukes i konkursprediksjonsmodeller. De testet både 50/50-fordeling, 80/20-fordeling med flest ikke-konkurser og 90/10-fordeling med flest ikke-konkurser. De kom frem til at 50/50-fordeling ga de beste prediksjonene. Dette er også i henhold til funnene i Wilson og Sharda (1994). Pelja og Wahlstrøm (2021) har balansert datasettet slik at det består av like mange konkursregnskaper som ikke-konkursregnskaper. Næss et al. (2017) benytter imidlertid en fordeling av datasettet hvor 1/3 av selskapene er klassifisert som konkurs og 2/3 som ikke-konkurs. Det finnes ingen standard for hvordan denne fordelingen skal være dersom man velger å vekte datasettet, og dermed finnes det mange variasjoner på hvordan datasettet balanseres. Ifølge Perez (2006) er det ingen forskningsbaserte generaliseringer for hvilken fordeling mellom de ulike klassene som er å foretrekke.

Ved å benytte den fordelingen som finnes naturlig i datagrunnlaget vil det oppstå et klassebalanseproblem som vil kunne forvrengte modellene (Liang et al., 2016), og i denne oppgaven er det derfor valgt å foreta en balansering av datasettet. En artikkel av Li og Sun (2012) taler også for å gjennomføre balansering mellom klassene, da resultatene her viser at modellen hadde bedre prediksjonsevne med et balansert datasett. Dette kan tyde på at balansering av datasettet vil være gunstig ved konkursprediksjon som benytter dyplæring.

Vektingen av klassene i denne oppgaven er basert på resultatene fra balanseringer som er gjort i den tidligere forskningen. Etter balanseringen bestod datasettet av like mange konkurs og ikke-konkursregnskaper, slik at det er ble en 50/50-fordeling mellom de to klassene. Siden antallet konkurser var mindre enn antallet ikke-konkurser, ble alle disse tatt med i modellene for å sikre så stort datagrunnlag som mulig. Figur 4 viser hvordan det deretter ble gjennomført undersampling for å redusere mengden ikke-konkurser. Undersampling ble benyttet da det, som nevnt i seksjon 3.4.3, ikke er ønskelig å ha flere identiske observasjoner i datasettet slik som man ville fått ved bruk av oversampling.

Det finnes ulike strategier for å velge ut observasjoner fra den største klassen, slik at de to klassene får samme størrelse. Pelja og Wahlstrøm (2021) benytter eksempelvis *matching*, noe som fører til at man får relativt like observasjoner i de ulike klassene basert på utvalgte kriterier. Denne måten å velge ut observasjoner på er særlig relevant dersom man skal gjøre undersøkelser knyttet til spesifiserte dimensjoner i datasettet, som for eksempel størrelse eller bransje. Dette er ikke tilfellet for denne oppgaven, ettersom oppgaven har et mer overordnet syn på konkursprediksjon. Ikke-konkurs-selskapene ble derfor tilfeldig valgt ut blant alle observasjonene i klassen, noe som også er tilfellet i Næss et al. (2017). På denne måten vil det balanserte datagrunnlaget bestå av observasjoner på tvers av alle regnskapsårene for begge de to klassene.



Figur 4: Fordeling av konkurser før og etter undersampling

4.2 Estimering

I denne seksjonen vil forskningstilnærmingen i oppgaven gjennomgås. Det blir sett på valg av hyperparameterne for de nevrale nettverkene, samt andre valg som er gjort ved oppsett og kjøring av modellene.

4.2.1 Hyperparameteroptimering for nevrale nettverk

I dyplæringsmodeller er det mange hyperparametere som kan settes opp med ulike verdier. Det er ingen eksakt vitenskap å sette opp et nevralt nettverk, og det er derfor ingen regler for hvilke verdier de ulike hyperparameterne skal ha (Chollet, 2021, s. 145; Zhang et al., 1999). Verdiene på hyperparameterne kan ha stor innvirkning på resultatet av modellene. Hvordan disse burde stilles inn avhenger av type og mengde data, samt hvor kompleks modellen er. Det er derfor lurt å se på tidligere forskning for å få forslag til hvilke verdier hver hyperparameter burde settes til. Utover dette er modelloppsettet stor grad av prøving og feiling, og kan ifølge Zhang et al. (1999) i større grad sees på som kunst enn som forskning. Dette skiller nevrale nettverk fra mange andre modeller hvor det finnes et sett regler eller anbefalinger for hvordan modellen kan settes opp på best mulig måte.

For å ha gode verdier på disse hyperparameterne i modellene har det blitt gjennomført en hyperparameteroptimering. I denne oppgaven har det blitt benyttet en bayesiansk optimeringsstrategi ved bruk av BoTorch og AxClient-bibliotekene for å velge ut de beste hyperparameterne for hver modell (Meta Platforms, u.å.). AxClient er en utvidelse i Python som gjør denne bayesianske optimeringen automatisk. Denne typen optimering brukes siden det er mange mulige kombinasjoner av hyperparametere, noe som gjør at det ville vært svært tidkrevende å teste disse kombinasjonene manuelt (Russel & Norvig, 2021, s. 689-690). Det optimeres egne verdier for hvert enkelt datasett i oppgaven, da de beste verdiene på hyperparameterne kan variere for ulike datasett. Ved optimeringen av hyperparameterne er det gjennomgått 60 iterasjoner som undersøker hyperparameterens mulighetsrom. Intervallene som utgjør disse mulighetsrommene er satt basert på innledende modellkjøringer og tidligere forskning.

Modellene i oppgaven bruker en *optimizer* som implementerer en Adam-algoritme for å justere de lineære kombinasjonene mellom nodene i hvert lag. Disse kombinasjonene kalles for vekter, og modellen justerer disse vektene i epoker gjennom hele treningsfasen. Det er *learning rate* som bestemmer hvor store justeringene skal være. I hyperparameteroptimeringen er mulighetsrommet til *learning rate* satt til å være mellom 0,00001 og 0,01. Disse verdiene er valgt da justeringene ikke skal være så store at man kun får tilfeldige resultater, men heller ikke så små at modellene blir sittende fast i lokale minimum (Chollet, 2021, s. 53).

Hovedmålet til dyplæringsmodellene er å minimere tapsfunksjonen, og vektene i modellen blir derfor justert basert på dette. Siden modellene løser klassifikasjonsproblemer med to klasser er det

binary_crossentropy som minimeres i denne oppgaven (Chollet, 2021, s. 164). Antallet epoker sier noe om hvor mange ganger modellen itererer for å finne en tilpasning som gir bedre prediksjon enn tidligere. Modellene i denne oppgaven trenes i maks 1000 epoker, i tillegg benyttes det *early stopping* som gjør at modellene stopper etter et gitt antall epoker uten forbedring. Kriteriet for å stoppe treningen er at valideringstapet ikke har hatt noen forbedring i løpet av fem påfølgende epoker.

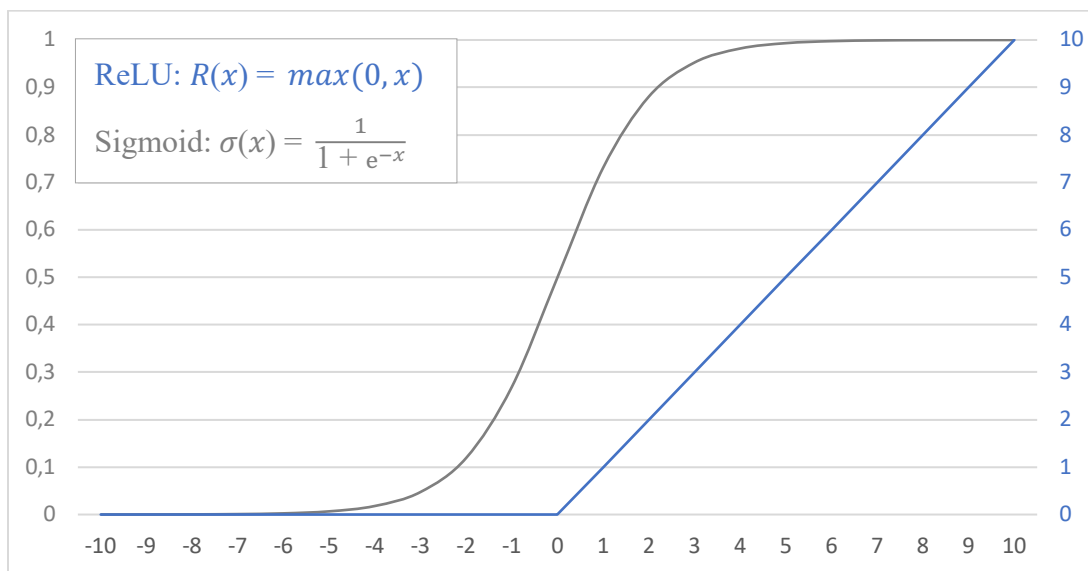
Batch size er betegnelsen på hvor mange observasjoner som skal inngå i hver epoke i treningen av modellene (Chollet, 2021, s. 91). Det er viktig å ha et tilstrekkelig antall observasjoner i hver syklus for at modellen skal kunne vurdere alle de mulige klassene og at variasjonen er stor nok til at modellen kan finne mønstre. En økning i *batch size* kan derfor føre til mer informative mønstre i dataene og mindre varians (Chollet, 2021, s. 139). I hyperparameteroptimeringen er *batch size* satt til å kunne ha verdiene 32, 64, 128, 512 og 1024. Disse verdiene er valgt ettersom det er vanlig praksis å benytte størrelser som er delelig på åtte i forskning som bruker maskinlæring.

Det er ingen fasit på hvilke verdier modellene skal ha for antall lag og noder (Chollet, 2021, s. 145). Modellene består av et input-lag, ett eller flere skjulte lag, og et output-lag, som vist i figur 2. I optimeringen av hyperparameterne er mulighetsrommet til antall lag i modellen satt til å være mellom 1 og 100. 100 lag ansees å være en fornuftig øvre grense for antall lag, siden den tidligere forskningen i seksjon 2.4 tyder på at modellen ikke vil trenge flere skjulte lag enn dette. Hvert lag i modellen har et antall noder som er en lineærkombinasjon av alle nodene i det foregående laget. I input-laget er antallet noder likt antallet variabler i datasettet som modellen skal trenes på (n). De skjulte lagene i modellen kan i prinsippet ha et hvilket som helst antall noder. Det kreves ofte mye eksperimentering for å komme frem til det mest gunstige antall noder, og hva som er det beste antallet noder er derfor vanskelig å bestemme før man har kjørt modellen (Zhang et al., 1999). Det finnes imidlertid noen tommelfingerregler for antallet noder i de skulte lagene, og det er vanlig å bruke et antall som henger sammen med antallet noder i inputlaget. Zhang et al. (1999) trekker frem antallene $n/2$, n , $n + 1$ og $2n + 1$ som forslag til antall noder i de skulte lagene, og det er dette som settes som mulige verdier i hyperparameteroptimeringen i denne oppgaven. Som nevnt i seksjon 2.4, har man som regel en eller to noder i det siste laget i modellen (Perez, 2006). I denne oppgaven har outputlaget kun én node, siden dette er et klassifiseringsproblem med to klasser (Chollet, 2021, s. 437).

Dropout er en av de mest effektive og mest populære metodene for regularisering av nevralt nettverk (Chollet, 2021, s. 150). *Dropout* implementeres i treningen av modellene ved at det i de skulte lagene innføres en sannsynlighet for at nodene ikke er aktive og dermed droppes av modellen. Sannsynligheten for at en node ikke er aktiv kalles for *dropout*-raten. En *dropout*-rate på 0,2 tilsier at 80 % av nodene i et lag er aktive. Dermed er det 20 % sannsynlighet for at hver node i det skjulte laget ikke er aktivt. Ved å fjerne en gitt andel noder i det nevralt nettverket vil modellen bli mindre kompleks, og det vil bli vanskeligere for modellen å fullstendig tilpasse seg treningssettet. Dette er, som beskrevet i seksjon 3.4.3, noe som ofte heller ikke er ønskelig. Ved å

sette *dropout*-raten for høyt vil man risikere å fjerne for mange noder i lagene, noe som gjør at modellen får lav prediksjonsevne. Det gjelder derfor å finne en balanse mellom modellens kompleksitet og prediksjonsevne. *Dropout*-raten settes ofte til en verdi mellom 0,2 og 0,5 (Chollet, 2021, s. 150). I innledende kjøring av modellene i denne oppgaven kommer det frem at resultatene blir enda bedre dersom en lavere *dropout*-rate tillates, og det er derfor ønskelig å sette den nedre grensen i mulighetsrommet til en liten verdi. Mulighetsrommet til *dropout*-raten i hyperparameteroptimeringen settes derfor til å være mellom 0,01 og 0,5.

Det benyttes to ulike aktiveringsfunksjoner i modellene; en *rectified linear unit* funksjon (ReLU) og en logistisk sigmoid funksjon. Aktiveringsfunksjonene anvendes for å begrense spekteret av verdier outputen fra modellene kan ha (Chollet, 2021, s. 115), og hvordan de to ulike funksjonene ser ut og hvilke verdier de kan gi vises i figur 5. ReLU er den mest populære aktiveringsfunksjonen i dyplæring (Chollet, 2021, s. 101), og brukes i alle de skjulte lagene i modellene. Funksjonen setter alle negative verdier til null. Den logistiske sigmoid aktiveringsfunksjonen brukes i outputlaget. Dette er vanlig i dyplæringsmodeller med to klasser, og naturlig å bruke ved et binært klassifiseringsproblem som konkursprediksjon (Chollet, 2021, s. 164; Perez, 2006). Denne aktiveringsfunksjonen gir *output* på en skala fra 0 til 1, og viser dermed sannsynligheten for at en gitt observasjon klassifiseres som 1 (Chollet, 2021, s. 100). I denne oppgaven betyr det at den viser sannsynligheten for at et selskap klassifiseres som konkurs basert på informasjonen fra variabelsettet.



Figur 5: Illustrasjon av de anvendte aktiveringsfunksjonene

For å klassifisere outputverdiene fra modellen inn i de to kategoriene konkurs og ikke-konkurs er det satt en *threshold*-verdi på 0,5. Siden *threshold*-verdien bestemmer hvordan observasjoner klassifiseres vil endringer i denne verdien kunne ha stor betydning for resultatet av modellene (Chollet, 2021, s. 166). Eksempelvis kan man bruke en lavere *threshold*-verdi for å veie opp for et

sterkt ubalansert datasett ved at man senker modellens terskel for å klassifisere observasjoner til den minste klassen. Dette er avveininger man må gjøre ut fra datasett og kontekst. I denne oppgaven er *threshold* på 0,5 benyttet, ettersom datasettene er balansert med like mange observasjoner i hver klasse. Denne *threshold*-verdien er brukt i tidligere forskning (Odom & Sharda, 1990). En *threshold* på 0,5 betyr at en observasjon klassifiseres som konkurs dersom den predikerte verdien fra outputlaget er større enn eller lik 0,5 og som ikke-konkurs dersom den predikerte verdien er mindre enn 0,5. Dette tolkes som at dersom modellen viser at det er over 50 % sannsynlighet for konkurs, så klassifiseres selskapet som konkurs.

De beste hyperparameterverdiene fra optimeringen for hvert datasett vises i tabell 4. Verdiene er rundet av til fire desimaler. For noen hyperparametere, som *learning rate* og antall skjulte lag, er det relativt mange like verdier for de ulike variabelsettene. Antall noder og *batch size* varierer imidlertid nokså mye. En av modellene får bare ett skjult lag, og selv om dette etter definisjonen i oppgaven ikke er dypplæring, brukes denne modellen slik den er. Dette grunnet at man i tidligere forskning ser at ett skjult lag er nok til at nevralt nettverk presterer godt (Næss et al., 2017; Odom & Sharda, 1990; Wilson & Sharda, 1994).

Tabell 4: Verdier fra hyperparameteroptimering

Variabelsett	Learning rate	Dropout rate	Skjulte lag	Noder		Batch size
Altman	0,0007	0,0526	3	11	$2n + 1$	256
SEBRA	0,0006	0,0113	3	15	n	32
Regnskap	0,0007	0,0816	9	73	$n + 1$	128
Regnskap-	0,0004	0,0384	3	24	$n/2$	512
Regnskap+ Altman	0,0004	0,0812	6	78	$n + 1$	256
Regnskap+ Altman-	0,0002	0,0364	1	141	$2n + 1$	128

4.2.2 Oppsett av modeller

De beste verdiene fra hyperparameteroptimeringen blir brukt til å trene oppgavens hovedmodeller. Hovedmodellene i denne oppgaven er forovermatete nevralt nettverk som er trent med *backpropagation*. *Backpropagation* er navnet på algoritmen som etter hver iterasjon oppdaterer vektene i lagene til modellen når den finner vektene som gir bedre resultater (Markham & Rakes, 1998). Som tidligere nevnt brukes tapsfunksjonen som et mål på hvor god modellen er. Algoritmen oppdaterer altså vektene når den finner en kombinasjon som gir lavere tap enn tidligere iterasjoner. De nevralt modellene kjøres med trenings-, validerings- og testsettene som er beskrevet i seksjon 3.4.3.

Ifølge Kirkos (2015) er det vanlig praksis å sammenlikne modeller på tvers av ulike klassifikasjonsmetoder. I denne oppgaven blir logistiske regresjonsmodeller benyttet som referansemodeller for modellene med dyplæring. Shumway (2001) omtaler *discrete hazard model* som en god sammenligningsmodell. Dette er en type logistisk regresjonsmodell som predikerer konkurs ut fra et perspektiv på flere år. Det gir bedre prediksjoner enn tradisjonelle statiske modeller siden modellen tar hensyn til et selskaps utvikling over tid (Paraschiv et al., 2021). I denne oppgaven er det imidlertid ikke inkludert noen tidstrend i datasettene, noe som gjør at de logistiske regresjonsmodellene i denne oppgaven er statiske, og dermed ikke kan analysere utviklingen i et selskap over tid. Dette gjøres for at det skal være direkte sammenlignbart med de nevrale nettverkene, hvor det heller ikke er inkludert noen tidstrend.

De logistiske regresjonene benytter det samme datagrunnlaget de nevrale nettverkene. Siden modellene med logistisk regresjon ikke trener modellene på samme måte som de nevrale nettverkene, består treningssettet til de logistiske regresjonene av både trenings- og valideringssettet fra seksjon 3.4.3. Testsettet i de logistiske regresjonene er det samme som i de nevrale nettverkene, og består av 10 % av det balanserte datagrunnlaget.

Logistiske regresjoner kan kun ta hensyn til lineære sammenhenger, og forutsetter derfor at observasjonene må være uavhengige av hverandre og at det ikke eksisterer spuriøse sammenhenger. De lineære sammenhengene gjør at avstanden mellom input og output i en logistisk regresjonsmodell er kort, da outputverdien består av den aggregerte summen av de ulike inputvariablene multiplisert med en enkelt vekting (Russel & Norvig, 2021, s. 801). En logistisk regresjon kan dermed sees på som et nevralt nettverk uten skjulte lag (Aas et al., 1999).

Oppgaven har dermed seks variabelsett (**Altman**, **SEBRA**, **Regnskap**, **Regnskap-**, **Regnskap+Altman**, **Regnskap+Altman-**) som testes med to ulike metoder (nevrale nettverk og logistisk regresjon). Dette resulterer i totalt tolv ulike modeller.

4.2.3 Tidshorisont for konkurs

Modellen predikerer om et selskap går konkurs med en horisont på ett år. Valget av konkurshorisont er en avveining mellom det å ha god treffprosent med kort konkurshorisont og det å kunne se tegn til konkurs tidlig. Det å ha en lengre tidshorisont gjør at man har mulighet til å kunne oppdage konkurser tidligere, men det kan gå på bekostning av modellens treffprosent (Altman et al., 1977). Det å ha kort konkurshorisont med god treffprosent, gjør at det kan bli vanskeligere å se hvilke selskaper som går konkurs på lengre sikt. Ett års konkurshorisont er mye brukt i tidligere forskning (Hillegeist et al., 2004; Tian & Yu, 2017; Tian et al., 2015). Det brukes også av Altman et al. (1977), Kainth og Wahlstrøm (2021) og Paraschiv et al. (2021) som arbeider med samme datagrunnlag. Siden regnskapene med verdien 1 for den avhengige variabelen er det siste regnskapet til et selskap, impliserer dette at det ikke leveres flere regnskaper for det gitte

selskapet. Dermed måles konkurser i modellen ut ifra at selskapet i løpet av det neste året slutter å levere regnskap, derav ett års horisont på konkursprediksjonen.

4.3 Evaluering

Etter at modellene er estimert må de evalueres for å sikre at de er egnet til konkursprediksjon. Det finnes flere ulike metoder for å evaluere modellens prestasjon, og videre gjennomgå evalueringmålene som benyttes for å teste prediksjonevnen til modellene i denne oppgaven.

Confusion matrix

En *confusion matrix* viser hvilke typer feil modellen har gjort. Det er en matrise hvor de predikerte output-verdiene fra modellen sammenliknes med de faktiske verdiene. Tabell 5 viser en generell *confusion matrix* med ulike typer feil. *True negative* (TN) og *true positive* (TP) er de tilfellene hvor modellen har predikert henholdsvis ikke-konkurs og konkurs korrekt. *False negative* (FN) og *false positive* (FP) er tilsvarende for de observasjonene som er klassifisert feil. Ved å bruke en *confusion matrix* kan man ta hensyn til at det er knyttet ulike kostnader til ulike typer feil. Dette er som tidligere nevnt noe som er relevant å tenke på når man arbeider med konkursprediksjon. Det skiller ofte mellom type I-feil og type II-feil, hvor type I og type II viser til henholdsvis FP og FN. For prediksjonen i denne oppgaven er det type II-feil som er mest alvorlig, da kostandene knyttet til å feilaktig predikere et selskap som ikke-konkurs er større enn å feilaktig predikere et selskap som konkurs (Altman et al., 1977; Stein, 2005).

Tabell 5: Confusion matrix med oversikt over ulike typer feil

Predikert verdi \ Faktisk verdi	Ikke-konkurs (0)	Konkurs (1)
Ikke-konkurs (0)	True negative (TN)	False positive (FP)
Konkurs (1)	False negative (FN)	True positive (TP)

Treffprosent

Treffprosent er et mål på hvor mange prosent av observasjonene modellen klassifiserer korrekt. Dette evalueringsmålet er enkelt å forstå, noe som også gjør det lett å tolke. Treffprosent er derfor mye benyttet i tidligere forskning (Kirkos, 2015; Næss et al., 2017). Dette er et godt mål på prestasjonen til modellen, men det tar ikke hensyn til hvilke typer feil modellen gjør (Næss et al., 2017). Det vil si at man ikke får noe informasjon om hvorvidt modellen gjør mange feil for en

bestemt klasse. Treffprosent viser antallet riktige prediksjoner delt på det totale antallet, og kan regnes ut med følgende formel:

$$\text{Treffprosent} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

Precision

Precision viser hvor mange prosent av de observasjonene som klassifiseres som konkurs som faktisk er konkurser, og kan dermed tolkes som et evalueringsmål som sier noe om kvaliteten på prediksjonene. Følgende formel brukes for å regne ut modellenes *precision*:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Recall

Et annet evalueringsmål som brukes for å evaluere modellene i denne oppgaven er *recall*. *Recall* omtales også ofte som sensitivitet og *true positive rate*, da det viser andelen korrekte konkursklassifiseringer av det totale antallet konkurser i datasettet. Evalueringsmålet sier altså noe om hvor stor andel av de faktiske konkursene modellen fanger opp, og kan dermed benyttes som mål for kvantiteten til klassifiseringene. Dette evalueringsmålet brukes siden det innenfor konkursprediksjon er knyttet en stor kostnad til å ikke oppdage konkurser. Det er imidlertid viktig å være oppmerksom på at en høy score på dette evalueringsmålet har lite verdi dersom alle observasjoner i modellen predikeres som konkurs. Formelen for *recall* kan skrives som:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

F1-score

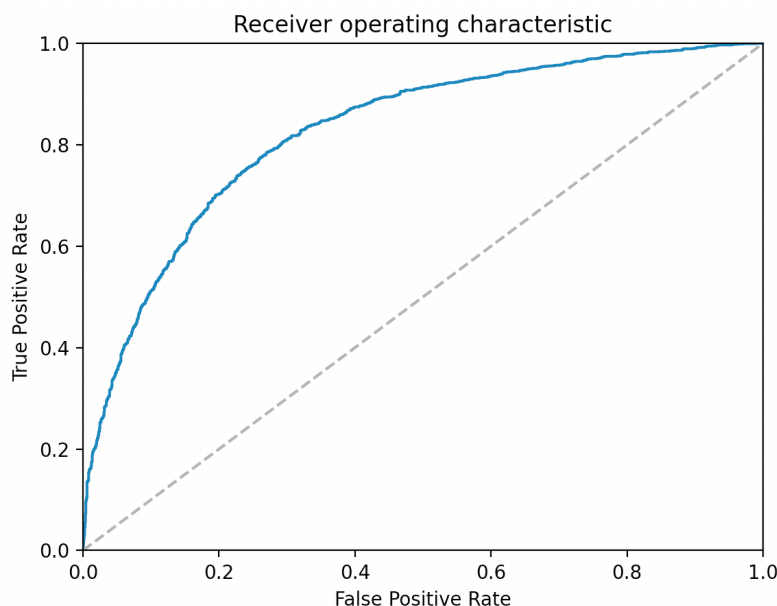
F1-score kombinerer *precision* og *recall* ved å finne det harmoniske gjennomsnittet av disse to. Dette evalueringsmålet kan for eksempel brukes til å skille to modeller hvor den ene har best *recall*, mens den andre har best *precision*. Formelen for F1-score kan skrives slik:

$$\text{F1} = \frac{2 * (\text{Precision} * \text{Recall})}{\text{Precision} + \text{Recall}}$$

ROC-kurve og AUC-verdi

Receiver operating characteristic (ROC)-kurven viser hvordan modellen klassifiserer observasjonene ved bruk av ulike *threshold*-verdier. En fremstilling av et ROC-diagram vises i figur 6. ROC-diagrammet viser modellens *true positive rate* på y-aksen og *false positive rate* ($\frac{\text{FP}}{\text{TP} + \text{FN}}$) på x-aksen. Ved å variere *threshold*-verdien til en modell vil man havne på ulike steder langs ROC-kurven. Dette tilsier at modellen vil ha forskjellige relative ytelser basert på hvilken

threshold som velges (Stein, 2005). I denne oppgaven benyttes det, som nevnt i seksjon 4.2.1, en *threshold* på 0,5. Ved å benytte andre *threshold*-verdier ville modellens inndeling i de to klassene blitt annerledes, og hvordan denne fordelingen endres ved varierende *threshold*-verdier kommer frem i ROC-diagrammet. De ulike *threshold*-verdiene vises imidlertid ikke direkte i diagrammet (James et al., 2021, s. 151).



Figur 6: Eksempel på ROC-diagram

Arealet under ROC-kurven kalles for *Area Under the Curve* (AUC), og er et mål som kan brukes for å evaluere og sammenlikne modeller. AUC-verdi er et mål som passer godt til analyser av konkursprediksjonsmodeller, ettersom det tar høyde for både kostnader ved feilklassifisering og skjeve fordelinger mellom de ulike klassene ved å ta hensyn til alle mulige *threshold*-verdier (Huang & Ling, 2005; James et al., 2021, s. 151). Dette evalueringsmålet er også i mye benyttet i den tidligere forskningen på konkursprediksjonsmodeller (Bernhardsen & Larsen, 2007; Kainth & Wahlstrøm, 2021; Næss et al., 2017; Tian & Yu, 2017; Tian et al., 2015). AUC-verdiene ligger mellom 0,5 og 1, og viser til modellenes forklaringskraft. Dersom en modell har AUC-verdi lik 0,5, vil ROC-kurven være lik den gråstiplede linjen i figur 6. Dette tilsier at modellen ikke er noe bedre enn tilfeldig gjetting. Hvis grafen ligger langt opp til venstre i figuren er forklaringskraften til modellen god. AUC-verdier mellom 0,5 og 0,7 ansees å være dårlige, verdier fra og med 0,7 til 0,8 ansees å være akseptable, verdier fra og med 0,8 til 0,9 ansees å være utmerket, mens verdier større enn eller lik 0,9 regnes som enestående (Hosmer et al., 2013, s. 177).

SHAP

For å visualisere påvirkningen hver variabel har på prediksjonene til modellen benyttes *Shapley Additive Explanations* (SHAP). SHAP har særlig stor nytte når man jobber med nevrale nettverk, siden dette som tidligere nevnt er en *black-box*-modell. Fra et økonomisk perspektiv er det viktig

å være klar over hvilke regnskapsposter og nøkkeltall som kjennetegner et selskap som er i ferd med å gå konkurs. Ved å benytte SHAP i denne oppgaven kan det hentes ut informasjon om hvilke variabler som er de viktigste driverne for konkurs i hver modell. Shapley-verdier er en måte å kvantifisere bidraget til hver enkelt variabel i modellen (Ancona et al., 2019). Det vil si hvor mye en gitt variabel påvirker utfallet til modellen. I et klassifiseringsproblem med to klasser vil det si hvor mye hver variabel marginalt trekker en observasjon mot 0 eller 1, i dette tilfellet konkurs eller ikke-konkurs. Det er Shapley-verdiene som bestemmer rekkefølgen på variablene i SHAP, og i denne oppgaven vil det bli lagt mest vekt på å analysere variablene med høyest Shapley-verdier.

McNemars test

McNemars test benyttes for å undersøke om det er signifikante forskjeller mellom modellene i oppgaven (McNemar, 1947). Dette er en statistisk test som utføres ved bruk av hypotesetesting. Nullhypotesen hevder at det ikke er noen signifikant forskjell mellom modellenes prestasjoner, mens den alternative hypotesen hevder at det er en signifikant forskjell i prediksjonsevnen til modellene. Dette er en metode som er lite brukt i den tidligere forskningen (Kirkos, 2015), men den benyttes blant annet i Næss et al. (2017) for å fastslå hvilken metode som har best prediksjonsevne. Testen kan kun benyttes til å avgjøre om det er signifikant forskjell eller ikke, da den ikke sier noe om hvor stor forskjellen mellom de ulike modellene er. I denne oppgaven blir den modellen som ansees som best basert på de andre evalueringskriteriene vurdert opp mot de resterende modellene som er laget med samme metode og mot alle modellene som er laget med den andre metoden.

4.4 Repliserbarhet

For at denne oppgaven skal være repliserbar er fremgangsmåten i analysen nøye beskrevet i kapittel 3 og 4. Det er også tatt forholdsregler i koden som benyttes med tanke på studiets repliserbarhet. Koden som er benyttet i studien kan oppgis på forespørsel. Datafilene kan dermed kjøres av andre og gi samme resultat. Dette bidrar til å sikre reliabilitet i oppgaven, ved at andre forskere kan gjenta studien med samme forskningstilnærming og få like resultater. Koden er videre skrevet slik at den henter ut de datafilene som skal brukes, og det er satt en fast verdi for *random seeds* for å sikre at nummergeneratorene i Python får samme verdier i hver kjøring ved randomisering. Datamaskinene og programvaren som er benyttet til å kjøre modellene i denne oppgaven er beskrevet i appendiks A6.

5 Resultater

I dette kapittelet presenteres resultatene fra de ulike modellene. Først presenteres resultatene fra referansem modellene laget med logistisk regresjon. Deretter presenteres resultatene fra de nevrale nettverkene, samt tilhørende SHAP-analyser. Kapittelet avsluttes med resultatene fra McNemars test for å undersøke om det er signifikante forskjeller mellom modellene.

5.1 Resultater fra logistisk regresjon

I denne seksjonen presenteres resultatene fra referansem modellene som ble laget med logistisk regresjon. Resultatene i tabell 6 viser at modellene med regnskapsvariabler tenderer til å gjøre det bedre enn modellene med **Altman** og **SEBRA** som er basert på finansielle nøkkeltall. Unntaket her er **SEBRA** som har høyere treffprosent enn **Regnskap-** med 77,77 % mot 77,57 %. Dette er en veldig marginal forskjell. Modellen med **SEBRA** har videre høyest *precision* med 78,28 %. Denne modellen har likevel lavere F1-verdi enn modellene med regnskapsvariabler. Dette tyder på at *recall* for **SEBRA** trekker F1-verdien ned. Det er også relativt klar forskjell på *recall* mellom variabelsett basert på finansielle nøkkeltall og de som er basert på regnskapsvariabler. Her skiller det over ti prosentpoeng mellom **Altman**, som er den med lavest *recall*, og **Regnskap-** som er den med høyest *recall*.

Av de logistiske regresjonsmodellene er det modellene basert på **Regnskap+Altman** og **Regnskap+Altman-** som har høyest treffprosent og AUC-verdi, mens det er modellene med **Regnskap** og **Regnskap-** som har høyest *recall*. Det er små forskjeller mellom F1-verdiene til disse modellene, men det er modellen med **Regnskap** som har høyest verdi. Alle AUC-verdiene fra de logistiske regresjonsmodellene ligger mellom 0,7 til 0,8, og ansees derfor å være akseptable basert på Hosmer et al. (2013, s. 177) sin inndeling.

Tabell 6: Oppsummering av resultater fra logistiske regresjoner

Variabelsett	AUC	Treffprosent	Precision	Recall	F1
Altman	0,7550	75,50 %	76,60 %	73,61 %	75,08 %
SEBRA	0,7778	77,77 %	78,28 %	77,04 %	77,66 %
Regnskap	0,7823	78,25 %	75,33 %	84,18 %	79,51 %
Regnskap-	0,7755	77,57 %	74,10 %	84,95 %	79,15 %
Regnskap+Altman	0,7886	78,87 %	77,78 %	80,99 %	79,35 %
Regnskap+Altman-	0,7878	78,78 %	77,74 %	80,81 %	79,25 %

5.2 Resultater fra nevrale nettverk

I denne seksjonen blir modellene som er laget med nevrale nettverk presentert. Som vist i tabell 7 er det modellene som er basert på regnskapsposter som har høyest prestasjonsevne. Modellene

med **Altman** og **SEBRA** skiller seg ut med noe dårligere verdier for samtlige av evalueringsmålne som er brukt i denne oppgaven. Modellen basert på variabelsett **SEBRA** er imidlertid bedre enn modellen med **Altman**.

Blant modellene som er basert på ubehandlede regnskapsdata er det relativt små variasjoner. F1-verdiene til modellene tyder på at modellene med **Regnskap** og **Regnskap-** presterer noe bedre enn modellene som også inkluderer Altman-nøkkeltall. Modellen med **Regnskap+Altman** er marginalt bedre enn den med **Regnskap+Altman-** på alle evalueringsmålne.

I denne oppgaven gir **Regnskap** marginalt bedre prediksjoner enn **Regnskap-**. Modellen med variabelsett **Regnskap** har best treffprosent med 81,09 %, høyest AUC-verdi med 0,8909 og høyest verdi på *precision* med 80,02 %. Den har imidlertid ikke høyest *recall*. Det er det modellen med **Regnskap-** som har med en verdi på 83,47 %. Ut fra F1-verdiene, som kombinerer *recall* og *precision*, er modellen med variabelsett **Regnskap** likevel marginalt best. En mer detaljert fremstilling av modellenes prediksjoner finnes i *confusion matrix* i appendiks A3.

Ifølge Hosmer et al. (2013, s. 177) sine rangeringer av modeller basert på modellenes AUC-verdi ansees alle modellene som settes opp med nevrale nettverk å prestere utmerket, da alle har verdier som ligger mellom 0,8 og 0,9.

Tabell 7: Oppsummering av resultater fra nevrale nettverk

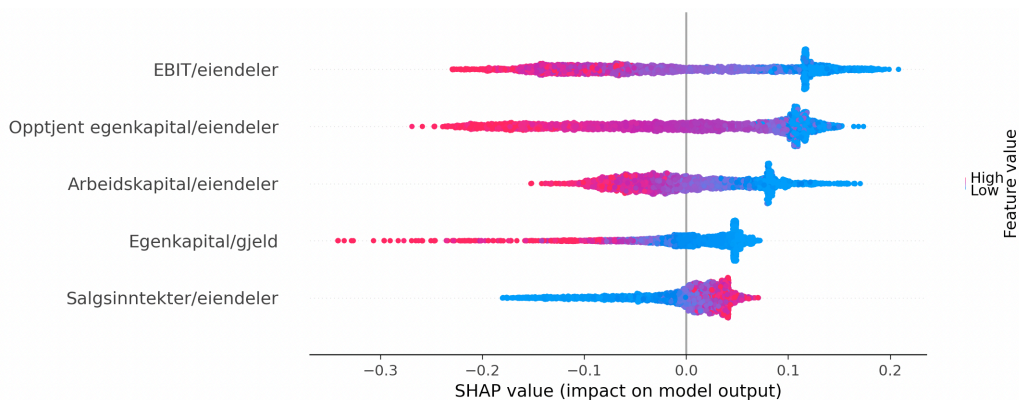
Variabelsett	AUC	Treffprosent	Precision	Recall	F1
Altman	0,8457	76,98 %	77,52 %	76,15 %	76,83 %
SEBRA	0,8703	79,05 %	79,88 %	77,80 %	78,83 %
Regnskap	0,8909	81,09 %	80,02 %	83,00 %	81,48 %
Regnskap-	0,8832	80,76 %	79,26 %	83,47 %	81,31 %
Regnskap+Altman	0,8896	80,65 %	79,31 %	83,06 %	81,14 %
Regnskap+Altman-	0,8858	79,96 %	78,62 %	82,47 %	80,50 %

5.3 SHAP-verdier

Ved å gjøre en SHAP-analyse av modellene gir det innsikt i hvilke variabler som påvirker modellens klassifikasjoner mest. Dette er interessant å se på for å vurdere de ulike variabelsettene opp mot hverandre. SHAP-analysene som ikke presenteres under, finnes i appendiks A5.

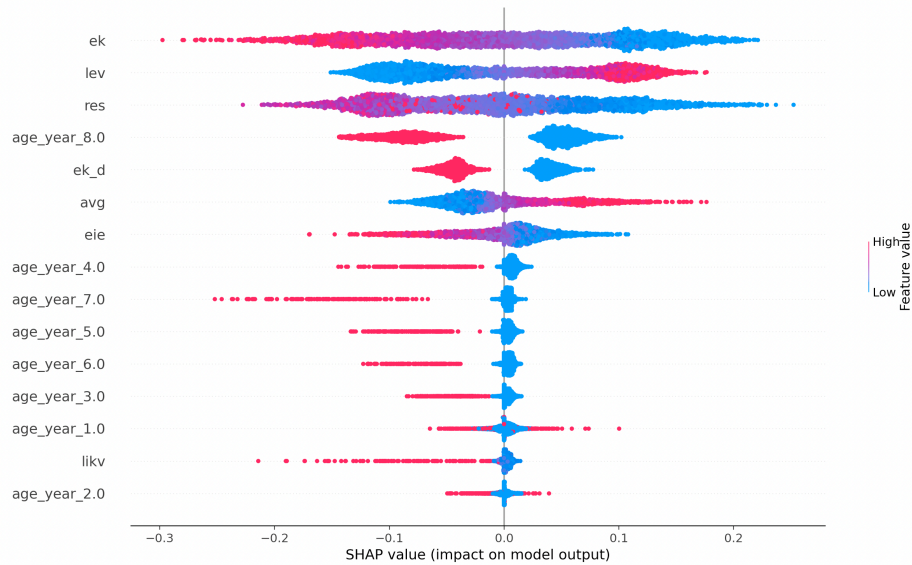
Figur 7 viser at i SHAP-analysen av modellen med **Altman** er det variabel X_3 (EBIT/eiendeler) og variabel X_2 (opptjent egenkapital/eiendeler) som er de to viktigste variablene. Disse gir begge økt sannsynlighet for konkurs ved lave verdier. For de fire variablene med størst påvirkning på modellen gir høyere verdier lavere sannsynlighet for konkurs.

For variabel X_5 (salgsinntekter/eiendeler) gir imidlertid høyere verdier større sannsynlighet for konkurs. Denne variabelen har minst påvirkning på modellen og har også et mindre tydelig skille mellom høye og lave verdier.



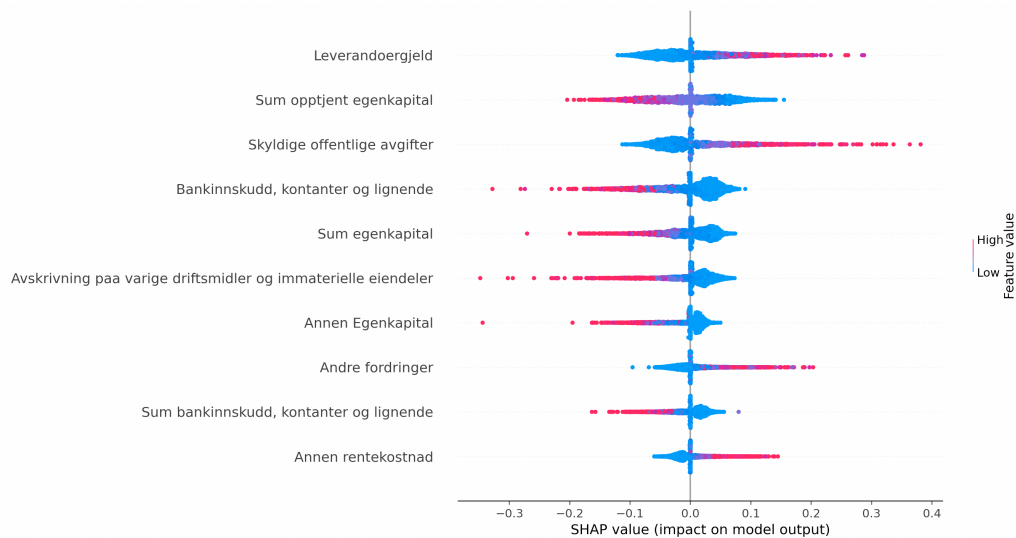
Figur 7: SHAP-analyse av variabelsett Altman

SHAP-analysen i figur 8 viser hvilke variabler som er de mest sentrale i modellen basert på variabelsett **SEBRA** trent med nevralt nettverk. Lave verdier av egenkapital og høye verdier av leverandørgjeld er to eksempler på forhold som trekker modellen i retning av å klassifisere et regnskap som konkursregnskap. Videre er det et tydelig skille i variabelen ek_d som angir om regnskapet gir positiv eller negativ egenkapital. Det er også tydelig at høye skyldige offentlige avgifter gir stor sjans for konkurs. Variabelen for resultat gir imidlertid et noe mer utydelig utslag på modellen. I denne variabelen er det ikke et like tydelig skille mellom høye og lave verdier som i mange av de andre variablene. Tendensen er likevel at høye resultater trekker mot lavere sannsynlighet for konkurs. Dummyene for alder gir et tydelig bilde av at yngre selskaper har større sjans for konkurs enn eldre selskaper. Variabelen $age_year_8.0$ er den variabelen med fjerde størst påvirkning på modellen. Denne har også et svært tydelig skille mellom selskaper som er over og under åtte år.



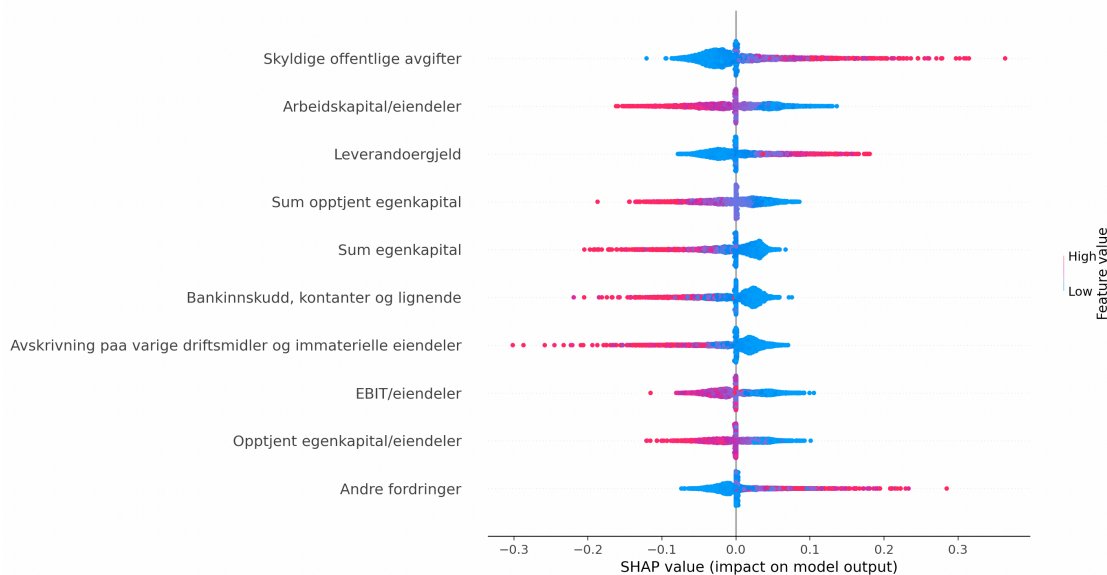
Figur 8: SHAP-analyse av variabelsett SEBRA

Figur 9 viser en SHAP-analyse av variabelsett **Regnskap** trent med nevralt nettverk, som er den modellen med best prediksjoner. Her er det postene leverandørgjeld, sum opptjent egenkapital og skyldige offentlige avgifter som er de mest innflytelsesrike variablene i modellen. Førstnevnte og sistnevnte gir høyere sannsynlighet for å klassifisere et selskap som konkurs ved høye verdier, mens sum opptjent egenkapital gir lavere sannsynlighet for konkurs ved høye verdier. Dette er forventet med bakgrunn i økonomisk teori. De øvrige variablene i figuren gir også utslag som er i tråd med forventningene.



Figur 9: SHAP-analyse av variabelsett Regnskap

I SHAP-analysen av den nevrale nettverksmodellen med **Regnskap+Altman**, vist i figur 10, er det mange av de samme variablene som har stor påvirkning på modellen som i modellen med **Regnskap**. Dette er blant annet skyldige offentlige avgifter og leverandørgjeld. I tillegg er det flere av Altman-nøkkeltallene som er sentrale variabler i denne modellen. Det er også flere tilfeller av de underliggende variablene som utgjør Altman-nøkkeltallene på listen over de mest innflytelsesrike variablene. Et eksempel på sistnevnte er sum opptjent egenkapital.



Figur 10: SHAP-analyse av variabelsett Regnskap+Altman

5.4 McNemars test

For å sammenligne modellene og se om forskjellene er signifikante benyttes McNemars test. **Regnskap** trent med nevrale nettverk er her brukt som referanse siden dette er den modellen som samlet sett presterte best. I tabell 8 er en oversikt over p-verdiene fra testen. Disse verdiene viser hvor mange prosent sannsynlig det er at en gitt modell er like god som referansmodellen. **Regnskap** trent med nevrale nettverk er ikke signifikant bedre enn **Regnskap+Altman** og **Regnskap+Altman-**, men den er signifikant bedre enn **Altman** og **SEBRA** på 1 %-nivå og bedre enn **Regnskap-** på 10 %-nivå. Denne modellen er også signifikant bedre enn alle modellene trent med logistisk regresjon, utenom modellene med **Regnskap+Altman** og **Regnskap+Altman-**.

Tabell 8: p-verdier fra McNemars test

	Nevrale nettverk	Logistisk regresjon
Altman	0,000***	0,000***
SEBRA	0,000***	0,000***
Regnskap	Ref.	0,000***
Regnskap-	0,087*	0,000***
Regnskap+Altman	0,196	0,702
Regnskap+Altman-	0,161	0,828

Resultater som er signifikant forskjellige fra referansemodellen (Ref.) på 10 %-nivå, er markert med *, på 5 %-nivå med ** og på 1 %-nivå med ***.

6 Drøfting

I dette kapittelet drøftes resultatene fra kapittel 5. Det trekkes frem funn og sammenhenger som anses å være relevante for problemstillingen.

6.1 Modeller med ubehandlede regnskapsdata og modeller med nøkkeltall

Resultatene i denne oppgaven tyder på at modeller laget med ubehandlede regnskapsdata er bedre enn modeller som er laget utelukkende med finansielle nøkkeltall. Modellene med **Altman** og **SEBRA** presterer dårligere enn modellene med variabelsett som inneholder en stor mengde rene regnskapsposter. En forklaring på dette kan være at disse modellene fanger opp momenter som ikke inngår i de få nøkkeltallene som er brukt i variabelsettene **Altman** og **SEBRA**.

I modellene som er basert på variabelsett med regnskapsposter kan modellen bruke svært mange ulike regnskapsposter til å finne kjennetegn på selskaper som holder på å gå konkurs. Et eksempel på en regnskapsvariabel som ikke inngår i modellene med **Altman** og **SEBRA** er avskrivning på varige driftsmidler og immaterielle eiendeler. I appendiks A5 vises det at SHAP-analysene til samtlige av modellene med regnskapsvariabler har denne avskrivningsposten blant de ti mest innflytelsesrike variablene i modellen. Det vises at høye verdier av denne variabelen trekker mot ikke-konkurs. På den ene siden kan dette skyldes at selskaper som nærmer seg konkurs vil forsøke å selge unna eiendeler og bruke regnskapsprinsipper som gir lave avskrivninger, ettersom dette kan få resultatene til å bedre seg på kort sikt. På den andre siden kan dette skyldes at større mer etablerte selskaper har større balanseførte verdier, noe som fører til høyere avskrivninger. Slike selskaper har som tidligere nevnt lavere sannsynlighet for konkurs.

Et annet eksempel på en post som har innvirkning på modellene med ubehandlede regnskapsdata er størrelsen på utbytte. Denne posten vises ikke i SHAP-analysene som presenteres i oppgaven, men den er rangert innenfor de 15 mest innflytelsesrike variablene for samtlige av disse modellene. Utbytte er en post som kan si mye om den økonomiske situasjonen i et selskap. Dersom et selskap deler ut penger til aksjonærene er det et sterkt tegn på at økonomien går bra. Det er også regulert hvorvidt det er lov å betale ut utbytte avhengig av hvor god økonomien i selskapet er (Aksjeloven, 1997, § 8-1). Utbytte er en post som ikke inngår i variabelsettene **Altman** og **SEBRA** i denne oppgaven. Utbytte var imidlertid med i en tidligere versjon av SEBRA-modellen (Eklund et al., 2001), ettersom utbytte ble ansett som et forhold som gjenspeilet hvorvidt eierne forventet fremtidig profitt. Bernhardsen og Larsen (2007) argumenterer for at utdeling av utbytte like gjerne kan skyldes skattetilpasning, og har derfor fjernet utbyttevariabelen i de nyere SEBRA-modellene. Hvorvidt det er riktig å inkludere denne variabelen kan derfor avhenge av hvordan man tolker den. I denne oppgaven vises det at den har en stor innvirkning på hvorvidt et selskap klassifiseres som

konkurs eller ikke, og det ansees derfor som riktig å ha den med da oppgavens formål er å forbedre konkursprediksjonsmodeller.

En annen potensiell forklaring på at modellene som består av variabelsett med ubehandlede regnskapsdata har bedre prediksjonskraft enn modellene med **Altman** og **SEBRA** er at disse modellene har en større mengde data å jobbe med. Datasettene med regnskapsposter er større siden det er like mange observasjoner for alle modellene, men flere variabler i variabelsettene med ubehandlede regnskapsdata. Dette er tilsvarende funnene i Næss et al. (2017) hvor de modellene med flest variabler gjør det best.

At det er en forskjell mellom modellene med regnskapsposter og de med **Altman** og **SEBRA** bekreftes også av McNemars test. Testen viser at det er signifikante forskjeller på 1 %-nivå for modeller både med nevrale nettverk og med logistisk regresjon for disse to variabelsettene, målt mot referansemodellen som er basert på **Regnskap** med nevrale nettverk. Dette underbygger funnet om at modeller med regnskapsvariabler gir bedre prediksjon enn modeller med **Altman** og **SEBRA**.

6.2 Modeller med **SEBRA** og **ALTMAN**

Modellen med **SEBRA** har marginalt bedre resultater enn den med **Altman** på alle evalueringsmålene ved bruk av nevrale nettverk. Ved bruk av logistisk regresjon er den også bedre på samtlige av evalueringsmålene som beskrevet i seksjon 4.3. Dette kan ha flere forklaringer, men det mest nærliggende er at variablene som er med i variabelsett **SEBRA** forklarer konkurser i norske selskaper på en bedre måte enn de variablene i **Altman** gjør. Et moment som taler for dette er at **SEBRA** er laget for norske selskaper av Norges Bank. Dette gjør at variabelsett **SEBRA** kan være mer tilpasset det norske markedet enn for eksempel variabelsett **Altman**.

En annen sentral forskjell mellom de to variabelsettene er at **SEBRA** har med dummyer for selskapets alder. Disse variablene er ikke inkludert i **Altman**, noe som bidrar til å forklare hvorfor modellene med **SEBRA** har bedre prediksjonsevne. Selskapets alder er statistisk sett viktig når det kommer til sannsynlighet for konkurs og dermed for mislighold av gjeld. Alders-dummyene i modellen med **SEBRA** har en klar trend ved at eldre selskaper har lavere sannsynlighet for konkurs, og at sannsynligheten øker jo yngre selskaper er. Som vist i seksjon 3.2 er det en overvekt av yngre selskaper som går konkurs. Dette er også støttet av tidligere forskning (Eklund et al., 2001; Paraschiv et al., 2021). I figur 8 vises det at dette er noe som fanges opp i modellen basert på variabelsett **SEBRA**. SHAP-analysen viser at særlig selskaper som er åtte år eller eldre har en klar reduksjon i sannsynlighet for konkurs. Dette er i tråd med hva man vil forvente ut fra økonomisk teori.

En annen forskjell på de to modellene er at variabelsett **SEBRA** har med en variabel for skyldige offentlige avgifter i prosent av totalkapitalen. Denne variabelen har høy påvirkning på modellen, og den er også utslagsgivende på flere av modellene i denne oppgaven. Dette er ikke overraskende ettersom et selskap risikerer strenge sanksjoner, som bøter og konkursbegjæring, dersom de offentlige avgiftene ikke blir betalt. Dette kan bidra til at modellene med **SEBRA** kan forklare noe som modellene med **Altman** muligens ikke fanger opp.

6.3 Sammenlikning av de fire variabelsettene med ubehandlede regnskapsdata

Alle de fire modellene som er basert på regnskapsvariabler har nokså lik prediksjonsevne, både med nevrale nettverk og logistisk regresjon. Dette gjelder for alle evalueringsmålene som er benyttet i denne oppgaven. Det er noen variasjoner mellom de to modellene som inkluderer Altman-nøkkeltallene og de to som ikke gjør det. Ved bruk av nevrale nettverk er det marginale forskjeller mellom de fire variabelsettene, men **Regnskap** og **Regnskap-** har marginalt høyere treffprosent og F1-verdi enn **Regnskap+Altman** og **Regnskap+Altman-**. Forskjellene er også små ved bruk av logistisk regresjon, men her skillet mellom variabelsettene med og uten Altman-nøkkeltallene litt mindre. Her har modellene med **Regnskap+Altman** og **Regnskap+Altman-** noe bedre resultater enn **Regnskap** og **Regnskap-** på treffprosent, *precision* og AUC-verdier. Forskjellene mellom disse er imidlertid litt større når man ser på *recall*, og på dette evalueringsmålet er det **Regnskap** og **Regnskap-** som gjør det best. F1-verdiene viser at det er modellen med **Regnskap** som gjør det best, etterfulgt av modellen med **Regnskap+Altman**. McNemars test viser at det er rimelig å anse **Regnskap** med nevrale nettverk som den beste modellen, men den kan ikke antas å være signifikant bedre enn noen av modellene med **Regnskap+Altman** og **Regnskap+Altman-**.

6.4 Modeller med **Regnskap** og **Regnskap-**

Variabelsett **Regnskap-** er identisk med **Regnskap**, med unntak av summeringspostene. Målet med å fjerne disse samlepostene var å undersøke om disse postene har en innvirkning på modellen. Resultatene i denne oppgaven viser at det for de ulike evalueringsmålene kun er marginale forskjeller på disse to modellene, men at modellen med summeringsposter er signifikant bedre på 10 %-nivå basert på McNemars test. Tre av de ti mest innflytelsesrike variablene i den nevrale nettverksmodellen med **Regnskap** er summeringsposter. Det at disse summepostene har så stor innvirkning på denne modellen taler for at modellene som inneholder disse postene vil kunne bli noe forskjellig fra modellene som ikke inkluderer disse. Resultatene gir ingen tydelige svar på om det er fornuftig å fjerne slike summeringsposter, men det er heller ingenting som tyder på at det er negativt å inkludere disse ettersom modellen basert på variabelsett **Regnskap** presterer best.

Modellene med **Regnskap** og **Regnskap-** har seks av de samme variablene innenfor de ti som har størst innflytelse på modellen. Leverandørgjeld, bankinnskudd, kontanter og liknende, skyldige offentlige avgifter, annen egenkapital, avskrivning på varige driftsmidler og immaterielle eiendeler og annen rentekostnad er dermed poster som har stor innvirkning på begge modellene. Dette er poster som det er naturlig å anta at kan si mye om et selskaps økonomiske situasjon. Høy kortsiktig gjeld kan tilsi at et selskap har utfordringer med likviditeten, noe som blant annet kan skyldes dårlig lønnsomhet. Slike utfordringer vil for eksempel kunne føre til at leverandørgjeld øker og bankinnskudd, kontanter og liknende reduseres. På lengre sikt vil dårlig lønnsomhet kunne føre til at poster som annen egenkapital reduseres. Det gir selskapet isolert sett en dårligere soliditet. Disse sammenhengene støttes av SHAP-analysene over de mest innflytelsesrike postene.

Av SHAP-analysene av disse modellene vises det også at mange av de underliggende postene som benyttes for å lage Altman-nøkkeltallene har høy innvirkning på disse modellene uten nøkkeltall. Dette tyder på at modellene ikke trenger Altman-nøkkeltallene for å finne sammenhengene i datagrunnlaget. *Sum opptjent egenkapital* og *sum egenkapital* er regnskapsposter som har stor innvirkning på modellen med **Regnskap**. Dette er poster som er fundamentale i konstruksjonen av Altman-nøkkeltallene. *Annen egenkapital* har høy innvirkning på modellen med **Regnskap-**, og dette kan tyde på at denne regnskapsposten utgjør en stor del av summeringspostene for egenkapital i variabelsett **Regnskap**. Det kan derfor virke som at modellen finner tilsvarende sammenhenger som det nøkkeltallene gir, selv om det ikke er de eksakt samme variablene som benyttes.

Sum omløpsmidler og *sum kortsiktig gjeld* er også viktige poster i utarbeidelsen av Altman-nøkkeltallene. SHAP-analysene viser at underposter for disse summeringspostene også har stor påvirkning på prediksjonen i de nevralt nettverksmodellene med **Regnskap** og **Regnskap-**. Variablene *bankinnskudd, kontanter og liknende* og *sum bankinnskudd, kontanter og liknende* er viktige i modellen med **Regnskap**, mens førstnevnte også har betydelig innvirkning i modellene basert på **Regnskap-**. Disse variablene er underposter i *sum omløpsmidler*. I både modellen med **Regnskap** og i modellen med **Regnskap-** har *leverandørgjeld* og *skyldige offentlige avgifter* høy innvirkning, og disse er underposter i *sum kortsiktig gjeld*. Dette er sterke tegn på at de nevralt nettverkene klarer å finne sammenhenger som likner på de finansielle nøkkeltallene på egenhånd, uten at de er eksplisitt oppgitt i variabelsettet.

6.5 Modeller med **Regnskap+Altman** og **Regnskap+Altman-**

Når det kommer til modellene basert på **Regnskap+Altman** og **Regnskap+Altman-** er det de underliggende regnskapspostene i Altman-nøkkeltallene som skiller disse fra hverandre, ved at de kun inngår i det førstnevnte variabelsettet. Som nevnt i 3.3.5 og 3.3.6 er disse variabelsettene inkludert for å undersøke hvor stor forklaringsstyrke Altman-nøkkeltallene har, og hva den isolerte effekten av disse nøkkeltallene er. De er også inkludert for å se på effekten av å ha med de underliggende variablene i nøkkeltallene, samt hvordan modellen påvirkes når også nøkkeltallene inkluderes. Resultatene viser at modellene med **Regnskap+Altman** presterer marginalt bedre enn modellene med **Regnskap+Altman-**. Dette kan skyldes at modellen med variabelsett **Regnskap+Altman** har noen flere variabler, og dermed mer treningsdata enn modellen uten disse underliggende regnskapspostene. En annen forklaring kan være at de åtte variablene som skiller disse variabelsettene gjør at variabelsettet **Regnskap+Altman** inneholder informasjon som gjør at prediksjoner med dette variabelsettet blir noe bedre.

SHAP-analysene av modellene basert på **Regnskap+Altman** og **Regnskap+Altman-** viser at disse modellene har høye Shapley-verdier for mange av de samme variablene. I begge de nevrale nettverksmodellene har skyldige offentlige avgifter størst innvirkning, etterfulgt av arbeidskapital/eiendeler og leverandørgjeld. Ettersom variabelen arbeidskapital/eiendeler har såpass stor innvirkning på modellenes prediksjoner tyder det på at dette nøkkeltallet i stor grad bidrar til modellenes forklaringskraft.

Variabelen opptjent egenkapital/eiendeler har høy innvirkning på modellen med **Regnskap+Altman-**. Dette tilsier at for modellen som ikke har de underliggende regnskapspostene i variabelsettet så bidrar Altman-nøkkeltallet som inneholder den samme informasjonen med relativt lik innvirkning på modellens prediksjon. I modellen med **Regnskap+Altman-** har variabelen annen egenkapital ganske høy innvirkning, og denne variabelen inneholder også mye av den samme informasjonen som nøkkeltallet. Dette kan tyde på at forklaringskraften til Altman-nøkkeltallet opptjent egenkapital/eiendeler isolert sett ikke er så veldig stor.

I det nevrale nettverket med **Regnskap+Altman** har sum opptjent egenkapital og sum egenkapital høy innvirkning. Dette viser at i modellen som har både nøkkeltallene og de underliggende regnskapspostene, så er det de underliggende postene som har høyest innvirkningskraft. Det tyder på at de underliggende regnskapspostene i Altman-nøkkeltallene har høyere innvirkningskraft på denne modellen, til tross for at Altman-nøkkeltallene også er med i variabelsettet. Dette bidrar til å underbygge at modellene på egenhånd klarer å fange opp de regnskapspostene som har høyest innvirkning på konkursprediksjon, og at dette er uavhengig av om det er flere variabler som inneholder mye av den samme informasjonen i variabelsettet.

En teori som kan forklare disse resultatene er at man med økende mengder datakraft i mindre grad trenger å bearbeide dataene i forkant modellestimeringen. Resultatene tyder på at de nevrale nettverksmodellene er noe bedre på dette enn modellene med logistisk regresjon. Det å ikke bearbeide datagrunnlaget i like stor grad som det gjøres i mye av den tidligere forskningen kan potensielt føre til at man får modeller som fanger opp andre og nye momenter, som man ut fra økonomisk teori ikke klarer å finne. Dette kan også gjøre modellene i stand til å utvikle seg i takt med økonomien, slik at modellene fortløpende vil kunne tilpasse seg endringer i selskapenes atferd.

6.6 Konsekvenser for banker

For banker som behandler lånesøknader er selskapets betalingsevne på kort sikt viktig. I SHAP-analysen av **Altman** kommer det frem at variabel X_3 (EBIT/eiendeler) er den variabelen med mest påvirkning på modellen. Dette er i henhold til resultatene i Altman (1968). Denne variabelen sier mye om selskapets lønnsomhet. I modellen med **Regnskap** er det stor overvekt av poster som kan ha innvirkning på et selskaps likviditet og soliditet. Lønnsomhet sier noe om et selskaps evne til å tjene penger, men for en bank er det kanskje enda viktigere at et selskap har penger til å betjene sine lån fortløpende. Likviditeten er da sentral på kort sikt, og soliditeten er viktig på litt lengre sikt. Disse to kategoriene av nøkkeltall sier mye om et selskaps betalingsevne, noe som gir banken forutsigbarhet når det kommer til selskapets betjening av renter og avdrag på lån. I praksis vil en bank naturligvis vurdere langt flere forhold enn de som blir drøftet i denne oppgaven, men mislighold av gjeld og konkurser henger likevel forholdsvis tett sammen.

6.7 Kostnader ved ulike typer feil

De ulike modellene i oppgaven har noe ulike egenskaper når det kommer til hvilken type feil de gjør. Som nevnt tidligere i oppgaven kan det være ulike kostnader knyttet til ulike typer feil. For en bank vil det eksempelvis være en større kostnad å tape hele eller deler av et utlån ved en konkurs, enn det vil være å ikke innvilge et lån og tape kunden til konkurrenten. De modellene som har høy *recall*, fanger opp mange av de faktiske konkursene. Alle modellene med regnskapsvariabler har god *recall* sammenliknet med modellene basert på **Altman** og **SEBRA**, noe som også vises i *confusion matrix* i appendiks A3. Det vil si at disse modellene vil være å foretrekke for en bank som anser det som veldig viktig å fange opp flest mulig konkurser.

Noe som kan påvirke forholdet mellom ulike typer feil er hvilken *threshold* som velges i modellen. I denne oppgaven er det, som beskrevet i 4.2.1, benyttet en *threshold* på 0,5. Ved å redusere denne ville man kunne få en høyere *true positive rate*, men samtidig en høyere *false positive rate*. Dette er en avveining når man skal lage modeller innenfor konkursprediksjon. Ved å se på ROC-diagrammene til modellene, vist i appendiks A4, kan man se hvordan dette forholdet endres med ulike *threshold*-verdier. Av ROC-diagrammene vises det at man på modellene med

regnskapsposter vil kunne øke *true positive rate*, og få en mindre økning i *false positive rate* enn det man kan med modellene basert på **Altman** og **SEBRA**.

Et eksempel for å demonstrere effekten av endret *threshold*-verdi er gjennomført på de nevrale nettverksmodellene med **Altman** og **Regnskap**. Med *threshold*-verdien som brukes i denne oppgaven får de nevrale nettverksmodellene med variabelsettene **Altman** og **Regnskap** henholdsvis 76,15 % og 83,00 % i *true positive rate*, og 22,08 % og 20,72 % i *false positive rate*. Dersom man ønsker en høyere *true positive rate* for modellene, vil *false positive rate* øke mer for modellen med **Altman** enn for modellen med **Regnskap**. Forholdet mellom de to feilratene vil være en avveining som avhenger av formålet med prediksjonen. Hvorvidt 0,5 er riktig *threshold* for disse modellene er en vurdering som må gjøres når modellene skal benyttes til et bestemt formål. Man kan gjennomføre en optimalisering av denne *threshold*-verdien, slik som gjøres av Hosmer (2013), men i en praktisk situasjon vil man vite mer om kostnadene ved ulike typer feil og på den måten kunne vurdere hvilken *threshold* som er mest fornuftig å bruke.

6.8 Modellenes kompleksitet

Som nevnt i seksjon 2.4 er det tidligere forskning som tyder på at mindre komplekse modeller fungerer best til konkurrerprediksjon med nevrale nettverk. I denne oppgaven er det derimot den modellen med flest skjulte lag som oppnår best resultater. Dette er som tidligere nevnt modellen som baserer seg på variabelsett **Regnskap**. Denne modellen er en av de mest komplekse modellene i oppgaven med 9 skjulte lag og 73 noder i hvert av disse lagene.

Et velkjent problem med komplekse modeller innenfor maskinlæring er at de blir overtilpasset. Det er derfor brukt *early stopping* i disse modellene, noe som gjør at modellene stopper å trene før de blir overtilpasset. Modellen med **Regnskap** er den modellen som stoppet etter færrest epoker. Det er ikke veldig store forskjeller knyttet til *early stopping* på tvers av modellene, selv om det er en tendens til at de mer komplekse modellene stopper tidligere enn de mindre komplekse. Modellene i oppgaven har ulik kompleksitet, så forskjeller i prediksjonsevne kan skyldes både ulikheter i oppsettet av modellene og hvilke variabler som er benyttet.

Under hyperparameteroptimeringen ble det, som tidligere nevnt, brukt et utvalg av antall noder som har blitt brukt i tidligere forskning. Det ble derfor inkludert både n og $n + 1$ antall noder. Siden disse alternativene er relativt like hverandre, kan det antas at det er av liten betydning hvilke av disse som brukes.

7 Konklusjon

I dette kapittelet fremlegges konkluderende bemerkninger fra oppgaven. Det presenteres også et kritisk blikk ved å gjennomgå oppgavens utfordringer og begrensninger. Kapittelet avsluttes med forslag til videre forskning innen konkursprediksjon ved bruk av nevrale nettverk og dyplæring.

7.1 Konkluderende bemerkninger

Resultatene som er presentert i denne oppgaven viser at modellen basert på variabelsett **Regnskap** med nevrale nettverk er signifikant bedre enn de modellene som kun inneholder finansielle nøkkeltall, både med logistisk regresjon og nevrale nettverk. Dette er basert på en McNemars test med 1 %-nivå. Dette tyder på at nevrale nettverk som benytter ubehandlede regnskapsdata i stor grad er i stand til å finne sammenhenger som man tidligere har satt sammen finansielle nøkkeltall for å forklare. Dette antyder at man med dagens datakraft kan lage modeller med alle variablene i et regnskap, og la modellen selv finne sammenhengene.

SHAP-analysene som er gjennomført viser hvilke variabler som i størst grad har påvirket prediksjonene til modellene. Resultatene fra disse analysene viser at det i modellene med **Regnskap** og **Regnskap-** i stor grad er regnskapsposter som det ut fra økonomisk teori er naturlig å anta at kan ha innvirkning på sannsynligheten for konkurs i et selskap som har høy påvirkning. Selv om modellene med rene regnskapsvariabler presterer bra, så viser SHAP-analysene at noen av nøkkeltallene fra variabelsett **Altman** har relativt stor innvirkning på modellene der disse nøkkeltallene er inkludert i tillegg til de ubehandlede regnskapsvariablene. Det vises imidlertid at regnskapsposter som inneholder likende informasjon som nøkkeltallene kan ha høyere innvirkning på modellen, noe som tyder på at disse nøkkeltallene ikke er betydelige for modellens prediksjonskraft. Basert på dette kan det virke som at man ikke behøver å ha med nøkkeltall i tillegg til de ubehandlede regnskapsdataene i variabelsettene, men modellenes prediksjonskraft blir ikke signifikant dårligere dersom nøkkeltallene inkluderes. Dersom man inkluderer nøkkeltallene kan det virke som at hvorvidt man har med de underliggende regnskapspostene disse nøkkeltallene bygger på heller ikke har så stor innvirkning på modellenes prediksjonskraft. Modellene presterer likevel marginalt bedre dersom de underliggende postene inkluderes. Resultatene viser også at dersom disse utelates, så tillegges nøkkeltallene større innvirkning på modellenes konkursprediksjon.

I tillegg viser resultatene at selv om den nevrale nettverksmodellen med variabelsett **Regnskap** ansees å være den beste modellen, så er det relativt små forskjeller i prediksjonsevnen mellom de ulike modellene som er basert på variabelsett som inneholder ubehandlede regnskapsdata. Det vises også at alle disse modellene gir jevnt over bedre resultater enn modeller med **SEBRA** og **Altman**. Uavhengig av hvilke av evalueringsmålene som er benyttet på modellene, gir også modellene basert på **SEBRA** marginalt bedre resultater enn modellene basert på **Altman**. Dette

viser en tendens til at modeller med flere variabler oppnår bedre resultater. SHAP-analysene av de seks nevralt nettverksmodellene viser også at påvirkningen hver variabel har på modellene er i tråd med det som er naturlig å anta ut fra økonomisk teori og tidligere forskning.

7.2 utfordringer og begrensninger

Modellene i denne oppgaven er laget med hovedfokus på å kunne gi et godt svar på oppgavens problemstilling: *I hvilken grad kan ubehandlede regnskapsdata forbedre konkursprediksjon med dyplæringsmodeller?* Dersom man skal lage slike modeller som skal brukes i praksis, må man se nærmere på hvordan det kan gjøres på en best mulig måte. Modellene i denne oppgaven gir informasjon om et selskaps regnskap klassifiseres som konkurs eller ikke, og det kunne vært spennende å se på et selskaps sannsynlighet for konkurs direkte. Dette ville vært gunstig i en praktisk situasjon, da det vil være forskjeller i den økonomiske situasjonen til selskaper utover om modellen klassifiserer et selskap til konkurs eller ikke med ett års konkurshorisont. Siden denne oppgaven har fokus på hvordan modellene klassifiserer og ser på ulike typer feil, er det mest hensiktsmessig å klassifisere konkurser som 0 og 1.

En utfordring med konkursprediksjonsmodeller er at det finnes mange ulike evalueringsmål. Denne oppgaven bruker treffprosent som evalueringsmål, og dette skyldes i stor grad at det i tidligere forskning er et stort fokus på dette målet. Treffprosent sier mye om prestasjonen til modellen, men er som nevnt i seksjon 4.4 ikke nødvendigvis det mest relevante evalueringsmålet i praksis. I denne oppgaven evalueres modellene derfor også med andre evalueringsmål som er utledet fra modellenes *confusion matrix*.

Det er veldig mange faktorer som kan ha innvirkning på hvordan et nevralt nettverk presterer. Det er derfor mye som kunne vært gjort annerledes når det kommer til oppsett av modellen. Eksempelvis kunne man økt *patience* på *early stopping*, brukt en annen *optimizer* eller bestemt de øvrige hyperparameterne på en annen måte. Grunnet oppgavens omfang er det begrenset hvor mange av disse parameterne som har blitt testet i denne oppgaven. Valget av modelloppsett vil kunne avhenge av hva man skal benytte modellen til i praksis, og i denne oppgaven er de nevralt nettverkene optimalisert for å oppnå lavest mulig tap.

Resultatene i oppgaven viser at enkelte modeller hadde bedre treffprosent på valideringssettet enn på treningssettet, som vist i appendiks A7. Dette virker litt ulogisk da modellene i utgangspunktet burde prestere best på data de er trent ved hjelp av, og ikke valideringssettet som ikke er inkludert i treningen. En mulig forklaring på dette kan være *dropout*-lagene som er benyttet i modellene. Disse fjerner et gitt antall tilfeldige noder i hvert lag under treningen av modellene. Dette kan gjøre det vanskeligere for modellene å tilpasse seg treningssettet, ettersom *dropout* brukes for å unngå overtilpasning. Enkelte av modellene i denne oppgaven er forsøkt trent også uten *dropout*. På disse kjøringene er treffprosenten bedre på treningssettet enn på valideringssettet, men treffprosenten på valideringssettet er dårligere enn på modellene som er trent med *dropout*. Dette forsterker teorien

om at *dropout* er grunnen til at prestasjonen på treningssettet er lavere enn på validerings- og testsettet. Denne effekten av *dropout*-lagene ansees som naturlig ettersom den reduserer sjansen for overtilpasning i modellene.

7.3 Forslag til videre forskning

Denne oppgaven konkluderer med at bruk av ubehandlede regnskapsdata kan brukes til konkursprediksjon med nevralt nettverk der finansielle nøkkeltall tidligere har vært brukt. Dette funnet gjør det interessant å forske videre på bruk av regnskapstall innenfor konkursprediksjon, for å se om tilsvarende konklusjoner kan trekkes med andre datagrunnlag og fremgangsmåter. Det å inkludere andre datatyper som for eksempel aksjekurser eller børsmeldinger kan også være interessant for å se om det kan forbedre dagens konkursprediksjonsmodeller. Det at regnskapsdata kun gjengir historiske hendelser er et annet moment som gjør det naturlig å lete etter mer fremtidsrettede datakilder. Eksempler på datakilder som kan si noe mer om konkurssannsynligheten til et selskap er budsjetter, strategiplaner og makrovariabler. Slik data er imidlertid ikke like lett tilgjengelig som det regnskapsdata er i dag.

Videre forskning på området oppfordres også til å se nærmere på hvordan man kan inkludere tidstrend for å se på utviklingen til et selskap over tid ved bruk av nevralt nettverk. Dette vil kunne si mer om hvordan et selskaps trend ser ut og kan potensielt gjøre prediksjon av konkurs bedre. For banker som skal vurdere en lånesøknad vil det være relevant å kjenne til sannsynlighet for konkurs med en lengre horisont enn ett år. Derfor kan det for banker være fordelaktig å ta i bruk andre analysemetoder i tillegg til ettårig konkursprediksjon. Eksempelvis vil det å gi selskaper en kredittscore kunne bidra til et mer nyansert bilde. For en bank vil en mer detaljert modell være nyttig i helhetsvurderingen av en lånesøknad. Det flere aspekter rundt et selskaps situasjon en bank må ta hensyn til ved behandling av søknader. Dette er forhold banken kvalitativt må vurdere. Eksempelvis ansees selskapets fremtidige kontantstrøm som særlig relevant. Det hadde vært spennende å se om man kan forbedre resultatene fra denne oppgaven ved å se videre på hvordan denne typen informasjon kan innlemmes i en dyplæringsmodell med ubehandlede regnskapsdata.

Opgaven bidrar til forskningen innenfor konkursprediksjon ved å se på hvilke variabler som egner seg best, samt hvordan dyplæringsmodeller kan brukes til å predikere konkurser. Ved å inkludere ubehandlede regnskapsdata i fremtidig forskning innenfor fagfeltet vil man kunne tilegne seg mer informasjon om hvordan slik data kan bidra til å forbedre konkursprediksjonsmodeller.

Referanseliste

- Aksjeloven. (1997). *Lov om aksjeselskaper* (LOV-1997-06-13-44). Lovdata.
<https://lovdata.no/dokument/NL/lov/1997-06-13-44>
- Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 23(4), 589-609. <https://doi.org/10.2307/2978933>
- Altman, E. I. (1993). *Corporate Financial Distress and Bankruptcy: A Complete Guide to Predicting & Avoiding Distress and Profiting from Bankruptcy* (2. utg.). Wiley.
- Altman, E. I. (2013). Predicting Financial Distress of Companies: Revisiting the Z-Score and ZETA® Models. I A. R. Bell, C. Brooks & M. Prokopczuk (Red.), *Handbook of Research Methods and Applications in Empirical Finance* (s. 428-456). Edward Elgar Publishing.
- Altman, E. I., Haldeman, R. G. & Narayanan, P. (1977). ZETA™ analysis A new model to identify bankruptcy risk of corporations. *Journal of Banking & Finance*, 1(1), 29-54.
[https://doi.org/10.1016/0378-4266\(77\)90017-6](https://doi.org/10.1016/0378-4266(77)90017-6)
- Ancona, M., Oztireli, C. & Gross, M. (2019). Explaining Deep Neural Networks with a Polynomial Time Algorithm for Shapley Values Approximation. I *Proceedings of the 36th International Conference on Machine Learning* (s. 272-281). PMLR.
- Begley, J., Ming, J. & Watts, S. (1996). Bankruptcy classification errors in the 1980s: An empirical analysis of Altman's and Ohlson's models. *Review of Accounting Studies*, 1(4), 267-284. <https://doi.org/10.1007/BF00570833>
- Bellovary, J. L., Giacomino, D. E. & Akers, M. D. (2007). A Review of Bankruptcy Prediction Studies: 1930 to Present. *Journal of Financial Education*, 33, 1-42.
<http://www.jstor.org/stable/41948574>
- Berg, D. (2007). Bankruptcy prediction by generalized additive models. *Applied Stochastic Models in Business and Industry*, 23(2), 129-143. <https://doi.org/10.1002/asmb.658>
- Bernhardsen, E. (2001). A Model of Bankruptcy Prediction. *Working papers from Norges Bank*, 10. <http://hdl.handle.net/10419/209799>
- Bernhardsen, E. & Larsen, K. (2007). Modelling av kredittrisiko i foretakssektoren – videreutvikling av SEBRA-modellen. *Penger og kreditt*, 35(2), 60-66.
- Chollet, F. (2021). *Deep Learning with Python* (2. utg.). Manning Publications Co.
- Eklund, T., Larsen, K. & Bernhardsen, E. (2001). Model for Analysing Credit Risk in the Enterprise Sector. *Penger og kreditt*, (3), 99-106. <http://hdl.handle.net/11250/2504900>
- Finanstilsynet. (2021). *Finansielt utsyn; Juni 2021* (6/2021).
<https://www.finanstilsynet.no/contentassets/de019705b5094a37ace5105e8b74b76d/finansielt-utsyn-juni-2021.pdf>
- Hillegeist, S. A., Keating, E. K., Cram, D. P. & Lundstedt, K. G. (2004). Assessing the Probability of Bankruptcy. *Review of Accounting Studies*, 9(1), 5-34.
<https://doi.org/10.1023/B:RAST.0000013627.90884.b7>
- Hosmer, D. W., Lemeshow, S. & Sturdivant, R. X. (2013). *Applied Logistic Regression* (3. utg.). John Wiley & Sons, Inc.
- Huang, J. & Ling, C. X. (2005). Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 17(3), 299-310.
<https://doi.org/10.1109/TKDE.2005.50>
- James, G., Witten, D., Hastie, T. & Tibshirani, R. (2021). *An Introduction to Statistical Learning* (2. utg.). Springer-Verlag New York Inc. <https://doi.org/10.1007/978-1-0716-1418-1>

- Kainth, A. & Wahlstrøm, R. R. (2021). Do IFRS Promote Transparency? Evidence from the Bankruptcy Prediction of Privately Held Swedish and Norwegian Companies. *Journal of Risk and Financial Management*, 14(3), 123. <https://doi.org/10.3390/jrfm14030123>
- Kirkos, E. (2015). Assessing methodologies for intelligent bankruptcy prediction. *Artificial Intelligence Review*, 43(1), 83-123. <https://doi.org/10.1007/s10462-012-9367-6>
- Konkursloven. (1986). *Lov om gjeldsforhandling og konkurs* (LOV-1984-06-08-58). Lovdata. <https://lovdata.no/dokument/NL/lov/1984-06-08-58>
- Konkursrådet. (2012, 24.01.12). *Innføring i konkurs*. Hentet 06.04.22 fra <https://www.konkursradet.no/innfoering-i-konkurs.305266.no.html>
- Langli, J. C. (2016). *Årsregnskapet* (10. utg.). Gyldendal Akademisk.
- Li, H. & Sun, J. (2012). Forecasting business failure: The use of nearest-neighbour support vectors and correcting imbalanced samples – Evidence from the Chinese hotel industry. *Tourism Management*, 33(3), 622-634. <https://doi.org/10.1016/j.tourman.2011.07.004>
- Liang, D., Lu, C.-C., Tsai, C.-F. & Shih, G.-A. (2016). Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. *European Journal of Operational Research*, 252(2), 561-572. <https://doi.org/10.1016/j.ejor.2016.01.012>
- Mansi, S. A., Maxwell, W. F. & Zhang, A. (2012). Bankruptcy Prediction Models and the Cost of Debt. *The Journal of Fixed Income*, 21(4), 25-42. <https://doi.org/10.3905/jfi.2012.21.4.025>
- Markham, I. S. & Rakes, T. R. (1998). The effect of sample size and variability of data on the comparative performance of artificial neural networks and regression. *Computers & Operations Research*, 25(4), 251-263. [https://doi.org/10.1016/S0305-0548\(97\)00074-9](https://doi.org/10.1016/S0305-0548(97)00074-9)
- McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12(2), 153-157. <https://doi.org/10.1007/BF02295996>
- Meta Platforms, I. (u.å.). *Bayesian Optimization*. Hentet 08.04.22 fra <https://ax.dev/docs/bayesopt.html>
- Nyitrai, T. & Virág, M. (2019). The effects of handling outliers on the performance of bankruptcy prediction models. *Socio-Economic Planning Sciences*, 67, 34-42. <https://doi.org/10.1016/j.seps.2018.08.004>
- Næss, A. B., Wahlstrøm, R. R., Helland, F. F. & Kjærland, F. (2017). Konkursprediksjon for norske selskaper - En sammenligning av regresjonsmodeller og maskinlæringsteknikker. I *Bred og spiss! NTNU Handelshøyskolen 50 år* (s. 313-330). Fagbokforlaget.
- Odom, M. D. & Sharda, R. (1990). A Neural Network Model for Bankruptcy Prediction. I *IJCNN International Joint Conference on Neural Networks* (s. 163-168). <https://doi.org/10.1109/IJCNN.1990.137710>
- Paraschiv, F., Schmid, M. & Wahlstrøm, R. R. (2021). Bankruptcy Prediction of Privately Held SMEs Using Feature Selection Methods. <https://doi.org/10.2139/ssrn.3911490>
- Pelja, I. & Wahlstrøm, R. R. (2021). Hvordan påvirker bedriftens størrelse predikering av konkurs? *Magma - Tidsskrift for økonomi og ledelse*, (7), 82-91.
- Perez, M. (2006). Artificial neural networks and bankruptcy forecasting: a state of the art. *Neural Computing & Applications*, 15(2), 154-163. <https://doi.org/10.1007/s00521-005-0022-x>
- Ravi Kumar, P. & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review. *European Journal of Operational Research*, 180(1), 1-28. <https://doi.org/10.1016/j.ejor.2006.08.043>

- Russel, S. & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach, Global Edition* (4. utg.). Pearson Education Limited.
- Shumway, T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *The Journal of Business*, 74(1), 101-124. <https://doi.org/10.1086/209665>
- Statistisk sentralbyrå. (2020). *Aksjeselskap er mest robuste*. Hentet 11.05.2022 fra <https://www.ssb.no/virksomheter-foretak-og-regnskap/artikler-og-publikasjoner/aksjeselskap-er-mest-robuste>
- Statistisk sentralbyrå. (2021). *06681: Nyetablerte foretak, etter overlevelse (K) 2001 - 2018* [Statistikk]. <https://www.ssb.no/statbank/table/06681>
- Statistisk sentralbyrå. (2022). *07164: Åpnede konkurser, etter næring (SN2007), organisasjonsform, omsetning, sysselsetting og etableringsår (F) 2006K1 - 2021K4* [Statistikk]. <https://www.ssb.no/statbank/table/07164/>
- Statistisk sentralbyrå. (u.å.). *Standard for næringsgruppering (SN)* (SN 2007). <https://www.ssb.no/klass/klassifikasjoner/6>
- Stein, R. M. (2005). The relationship between default prediction and lending profits: Integrating ROC analysis and loan pricing. *Journal of Banking & Finance*, 29(5), 1213-1236. <https://doi.org/10.1016/j.jbankfin.2004.04.008>
- Tian, S. & Yu, Y. (2017). Financial ratios and bankruptcy predictions: An international evidence. *International Review of Economics & Finance*, 51, 510-526. <https://doi.org/10.1016/j.iref.2017.07.025>
- Tian, S., Yu, Y. & Guo, H. (2015). Variable selection and corporate bankruptcy forecasts. *Journal of Banking & Finance*, 52, 89-100. <https://doi.org/10.1016/j.jbankfin.2014.12.003>
- Wahlstrøm, R. R. (2022). Financial statements of companies in Norway. *arXiv:2203.12842*. <https://doi.org/10.48550/arXiv.2203.12842>
- Wang, G., Ma, J. & Yang, S. (2014). An improved boosting based on feature selection for corporate bankruptcy prediction. *Expert Systems with Applications*, 41(5), 2353-2361. <https://doi.org/10.1016/j.eswa.2013.09.033>
- Wilson, R. L. & Sharda, R. (1994). Bankruptcy prediction using neural networks. *Decision Support Systems*, 11(5), 545-557. [https://doi.org/10.1016/0167-9236\(94\)90024-8](https://doi.org/10.1016/0167-9236(94)90024-8)
- Zhang, G., Y. Hu, M., Eddy Patuwo, B. & C. Indro, D. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research*, 116(1), 16-32. [https://doi.org/10.1016/S0377-2217\(98\)00051-4](https://doi.org/10.1016/S0377-2217(98)00051-4)
- Zhao, W. D. & Zhang, Y. D. (2014). Application Development of Student Score Analysis System for Secondary School Based on Standard Score. *Advanced Materials Research*, 989-994, 5062-5065. <https://doi.org/10.4028/www.scientific.net/AMR.989-994.5062>
- Zhao, Z., Xu, S., Kang, B. H., Kabir, M. M. J., Liu, Y. & Wasinger, R. (2015). Investigation and improvement of multi-layer perceptron neural networks for credit scoring. *Expert Systems with Applications*, 42(7), 3508-3516. <https://doi.org/10.1016/j.eswa.2014.12.006>
- Aas, K., Huseby, R. B. & Thune, M. (1999). *Data Mining - A Survey* (Report at the Norwegian Computing Center 942). Norsk Regnesentral.

Appendiks

A1: Variabler i variabelsett

A2: Utarbeidelse av Altman-nøkkeltall

A3: *Confusion matrix* for modellene

A4: ROC-diagrammer for modellene

A5: SHAP-analyser av modeller med nevralt nettverk

A6: Datamaskinspesifikasjoner og Python-bibliotek

A7: Treffprosent og tap på trenings- og valideringssett for nevralt nettverk

A1: Variabler i variabelsett

Variabler	Regnskap-	Regnskap	Regnskap +Altman	Regnskap +Altman-	Altman	SEBRA
Antall variabler	49	73	78	70	6	16
konkurs_dummy	X	X	X	X	X	X
Andre fordringer	X	X	X	X		
Annen driftsinntekt	X	X	X	X		
Annen driftskostnad	X	X	X	X		
Annen Egenkapital	X	X	X	X		
Annen finansinntekt	X	X	X	X		
Annen finanskostnad	X	X	X	X		
Annen innskutt egenkapital	X	X	X	X		
Annen kortsiktig gjeld	X	X	X	X		
Annen renteinntekt	X	X	X	X		
Annen rentekostnad	X	X	X	X		
Avskrivning på varige driftsmidler og immaterielle eiendeler	X	X	X	X		
Bankinnskudd kontanter og lignende	X	X	X	X		
Betalbar skatt	X	X	X	X		
Driftsløse inventar verktoy kontormaskiner og lignende	X	X	X	X		
Driftsresultat	X	X	X	X		
Gjeld til kredittinstitusjoner	X	X	X	X		
Investeringer i aksjer og andeler	X	X	X	X		
Konsernbidrag	X	X	X	X		
Konsernfordringer	X	X	X	X		
Kortsiktig konserngjeld	X	X	X	X		
Kundefordringer	X	X	X	X		
Leverandørgjeld	X	X	X	X		
Lønnskostnad	X	X	X	X		
Markedsbaserte aksjer	X	X	X	X		
Maskiner og anlegg	X	X	X	X		
Netto finans	X	X	X	X		
Øvrig langsiktig gjeld	X	X	X	X		
Ordinært resultat etter skattekostnad	X	X	X	X		
Ordinært resultat før skattekostnad	X	X	X			
Overføringer til/fra annen egenkapital	X	X	X	X		
Overkurs	X	X	X	X		
Overkursfond	X	X	X	X		

Variabler	Regnskap-	Regnskap	Regnskap +Altman	Regnskap +Altman-	Altman	SEBRA
Rentekostnad til foretak i samme konsern	X	X	X	X		
Salgsinntekt	X	X	X	X		
Selskapskapital	X	X	X	X		
Skattekostnad på ordinært resultat	X	X	X	X		
Skyldige offentlige avgifter	X	X	X	X		
Tomter bygninger og annen fast eiendom	X	X	X	X		
Totalresultat	X	X	X	X		
Udekket tap	X	X	X	X		
Utbytte	X	X	X	X		
Utsatt skatt	X	X	X	X		
Utsatt skattefordel	X	X	X	X		
Varekostnad	X	X	X	X		
Varer	X	X	X	X		
Årsresultat	X	X	X	X		
Årsresultat etter minoritetsinteresser	X	X	X	X		
Årsresultat før minoritetsinteresser	X	X	X	X		
Sum anleggsmidler		X	X	X		
Sum annen langsiktig gjeld		X	X	X		
Sum avsetninger for forpliktelse		X	X	X		
Sum bankinnskudd kontanter og lignende		X	X	X		
Sum egenkapital		X	X			
Sum egenkapital og gjeld		X	X	X		
Sum eiendeler		X	X			
Sum finansielle anleggsmidler		X	X	X		
Sum finansinntekter		X	X	X		
Sum finanskostnader		X	X	X		
Sum fordringer		X	X	X		
Sum gjeld		X	X			
Sum immaterielle eiendeler		X	X	X		
Sum innskutt egenkapital		X	X	X		
Sum inntekter		X	X			
Sum investeringer		X	X	X		
Sum kortsiktig gjeld		X	X			
Sum kostnader		X	X	X		
Sum langsiktig gjeld		X	X	X		

Variabler	Regnskap-	Regnskap	Regnskap +Altman	Regnskap +Altman-	Altman	SEBRA
Sum omløpsmidler		X	X			
Sum opptjent egenkapital		X	X			
Sum overføringer og disponeringer		X	X	X		
Sum varer		X	X	X		
Sum varige driftsmidler		X	X	X		
Arbeidskapital/eiendeler			X	X	X	
EBIT/eiendeler			X	X	X	
Egenkapital/gjeld			X	X	X	
Opptjent egenkapital/eiendeler			X	X	X	
Salgsinntekter/eiendeler			X	X	X	
avg						X
eie						X
ek						X
ek_d						X
lev						X
likv						X
res						X
age_year_1.0						X
age_year_2.0						X
age_year_3.0						X
age_year_4.0						X
age_year_5.0						X
age_year_6.0						X
age_year_7.0						X
age_year_8.0						X

A2: Utarbeidelse av Altman-nøkkeltall

Variablene som inngår i variabelsett **Altman** er basert på Altman (1968) sin definisjon av variablene i Z-score-modellen, og er i denne oppgaven definert som følgende:

	Altman (1968)	Utarbeidelse av nøkkeltall basert på regnskapsposter fra årsregnskapene	Variabelnavn i Altman
X ₁	$\frac{\text{working capital}}{\text{total assets}} = \frac{\text{arbeidskapital}}{\text{total kapital}}$	$\frac{\text{Sum omloepsmidler} - \text{Sum kortsiktig gjeld}}{\text{SUM EIENDELER}}$	$\frac{\text{arbeidskapital}}{\text{eiendeler}}$
X ₂	$\frac{\text{retained earnings}}{\text{total assets}} = \frac{\text{opptjent egenkapital}}{\text{total kapital}}$	$\frac{\text{Sum opptjent egenkapital}}{\text{SUM EIENDELER}}$	$\frac{\text{opptjent egenkapital}}{\text{eiendeler}}$
X ₃	$\frac{\text{earnings before interest and taxes}}{\text{total assets}} = \frac{\text{resultat før rentekostnad og skatt}}{\text{total kapital}}$	$\frac{\text{Ordinaert resultat før skattekostnad}}{\text{SUM EIENDELER}}$	$\frac{\text{EBIT}}{\text{eiendeler}}$
X ₄	$\frac{\text{total market equity}}{\text{book value of total debt}} = \frac{\text{egenkapitalens markedsverdi}}{\text{samlet gjeld}}$	$\frac{\text{Sum egenkapital}}{\text{Sum gjeld}}$	$\frac{\text{egenkapital}}{\text{gjeld}}$
X ₅	$\frac{\text{sales}}{\text{total assets}} = \frac{\text{salgsinntekter}}{\text{total kapital}}$	$\frac{\text{Sum inntekter}}{\text{SUM EIENDELER}}$	$\frac{\text{salgsinntekter}}{\text{eiendeler}}$

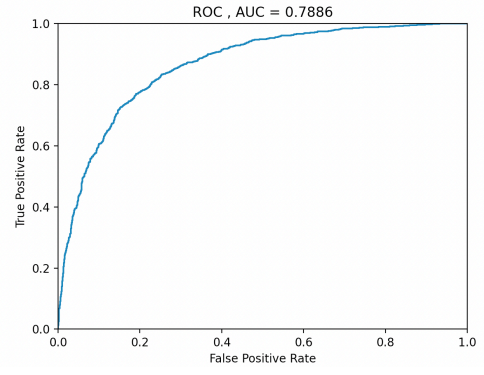
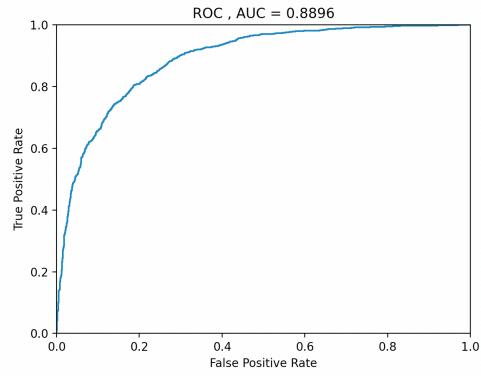
A3: *Confusion matrix* for modellene

		Predikert 0		Predikert 1	
		<i>Nevrale nettverk</i>	<i>Logistisk regresjon</i>	<i>Nevrale nettverk</i>	<i>Logistisk regresjon</i>
Faktisk 0	Altman	1311	1304	374	381
	SEBRA	1353	1323	332	362
	Regnskap	1334	1218	351	467
	Regnskap-	1315	1182	370	503
	Regnskap+Altman	1318	1293	367	392
	Regnskap+Altman-	1305	1293	380	392
Faktisk 1	Altman	404	447	1290	1247
	SEBRA	376	389	1318	1305
	Regnskap	288	268	1406	1426
	Regnskap-	280	255	1414	1439
	Regnskap+Altman	287	322	1407	1372
	Regnskap+Altman-	297	325	1397	1369

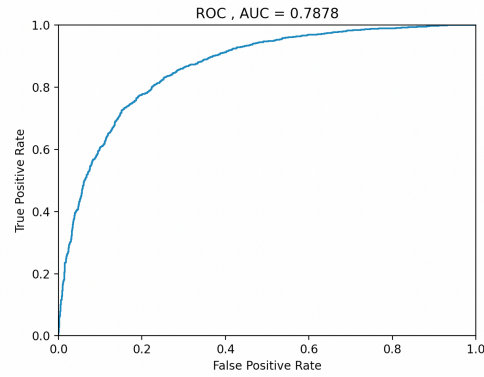
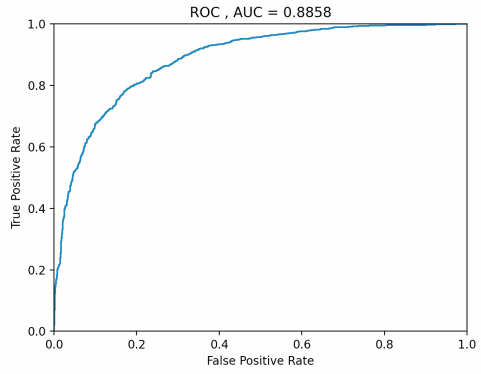
A4: ROC-diagrammer for modellene

Variabelsett	Nevrale nettverk	Logistisk regresjon
Altman	<p>ROC , AUC = 0.8457</p>	<p>ROC , AUC = 0.7550</p>
SEBRA	<p>ROC , AUC = 0.8703</p>	<p>ROC , AUC = 0.7778</p>
Regnskap	<p>ROC , AUC = 0.8909</p>	<p>ROC , AUC = 0.7823</p>
Regnskap-	<p>ROC , AUC = 0.8832</p>	<p>ROC , AUC = 0.7755</p>

Regnskap+
Altman

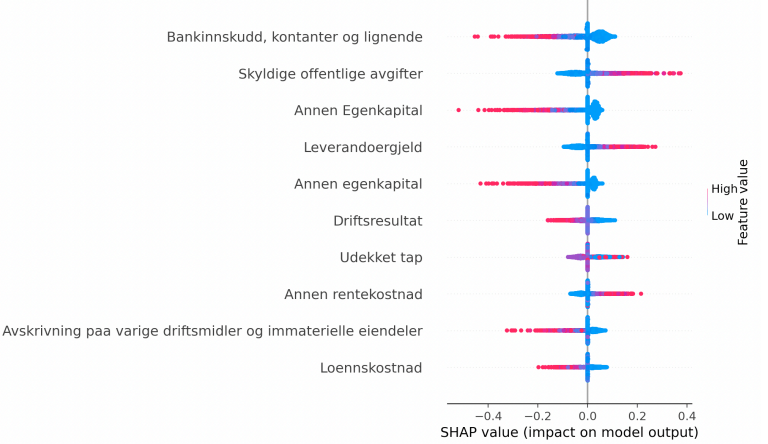
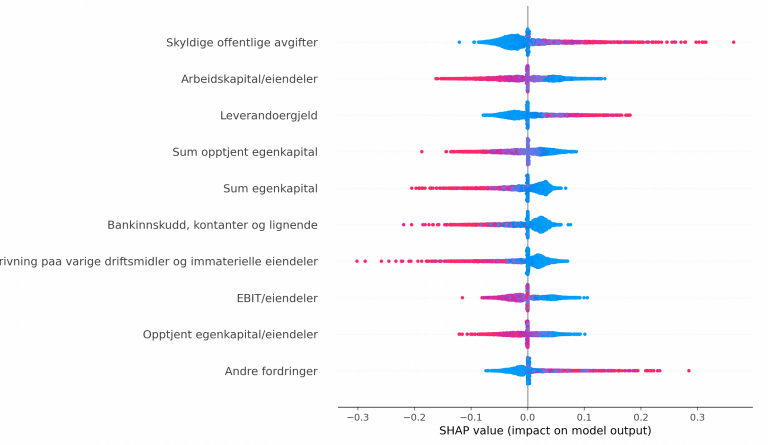
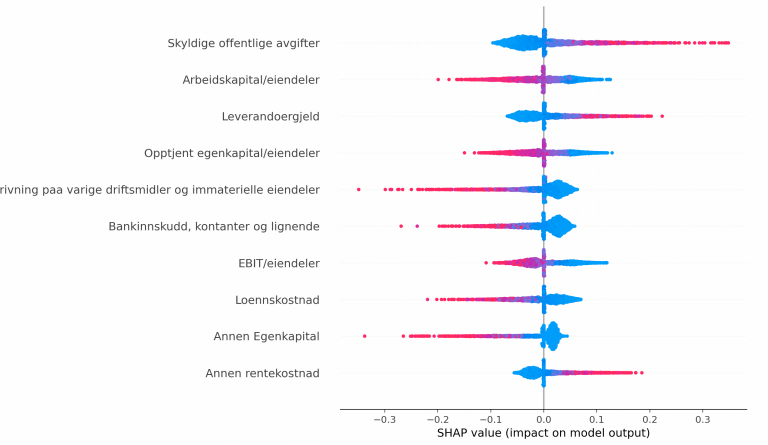


Regnskap+
Altman-



A5: SHAP-analyser av modeller med nevrale nettverk

Variabelsett	SHAP
Altman	<p>SHAP plot for the Altman model. The y-axis lists features: EBIT/eiendeler, Optjent egenkapital/eiendeler, Arbeidskapital/eiendeler, Egenkapital/gjeld, and Salgsinntekter/eiendeler. The x-axis is 'SHAP value (impact on model output)' ranging from -0.3 to 0.2. A vertical line is at 0.0. The plot shows the distribution of SHAP values for high (right) and low (left) feature values. EBIT/eiendeler and Optjent egenkapital/eiendeler show a strong positive impact for high values. Salgsinntekter/eiendeler shows a strong negative impact for high values.</p>
SEBRA	<p>SHAP plot for the SEBRA model. The y-axis lists features: ek, lev, res, age_year_8.0, ek_d, avg, eie, age_year_4.0, age_year_7.0, age_year_5.0, age_year_6.0, age_year_3.0, age_year_1.0, likv, and age_year_2.0. The x-axis is 'SHAP value (impact on model output)' ranging from -0.3 to 0.2. A vertical line is at 0.0. The plot shows the distribution of SHAP values for high (right) and low (left) feature values. Most features have a positive impact, with 'ek' and 'lev' showing the largest positive impacts for high values.</p>
Regnskap	<p>SHAP plot for the Regnskap model. The y-axis lists features: Leverandørgjeld, Sum optjent egenkapital, Skyldige offentlige avgifter, Bankinnskudd, kontanter og lignende, Sum egenkapital, Avskrivning paa varige driftsmidler og immaterielle eiendeler, Annen Egenkapital, Andre fordringer, Sum bankinnskudd, kontanter og lignende, and Annen rentekostnad. The x-axis is 'SHAP value (impact on model output)' ranging from -0.3 to 0.4. A vertical line is at 0.0. The plot shows the distribution of SHAP values for high (right) and low (left) feature values. Skyldige offentlige avgifter and Sum optjent egenkapital show significant positive impacts for high values.</p>

Variabelsett	SHAP
Regnskap-	 <p>SHAP plot for Regnskap-. The x-axis represents the SHAP value (impact on model output) ranging from -0.4 to 0.4. The y-axis lists features. The plot shows the distribution of SHAP values for each feature, with a vertical line at 0.0. Features with positive SHAP values (right of 0.0) include Skyldige offentlige avgifter, Annen egenkapital, Driftsresultat, Udekket tap, and Annen rentekostnad. Features with negative SHAP values (left of 0.0) include Bankinnskudd, kontanter og lignende, Annen Egenkapital, Leverandoergjeld, Avskrivning paa varige driftsmidler og immaterielle eiendeler, and Loennskostnad.</p>
Regnskap+ Altman	 <p>SHAP plot for Regnskap+ Altman. The x-axis represents the SHAP value (impact on model output) ranging from -0.3 to 0.3. The y-axis lists features. The plot shows the distribution of SHAP values for each feature, with a vertical line at 0.0. Features with positive SHAP values (right of 0.0) include Skyldige offentlige avgifter, Sum egenkapital, and Andre fordringer. Features with negative SHAP values (left of 0.0) include Arbeidskapital/eiendeler, Leverandoergjeld, Sum opptjent egenkapital, Bankinnskudd, kontanter og lignende, Avskrivning paa varige driftsmidler og immaterielle eiendeler, EBIT/eiendeler, and Opptjent egenkapital/eiendeler.</p>
Regnskap+ Altman-	 <p>SHAP plot for Regnskap+ Altman-. The x-axis represents the SHAP value (impact on model output) ranging from -0.3 to 0.3. The y-axis lists features. The plot shows the distribution of SHAP values for each feature, with a vertical line at 0.0. Features with positive SHAP values (right of 0.0) include Skyldige offentlige avgifter, Leverandoergjeld, Avskrivning paa varige driftsmidler og immaterielle eiendeler, Bankinnskudd, kontanter og lignende, EBIT/eiendeler, and Annen rentekostnad. Features with negative SHAP values (left of 0.0) include Arbeidskapital/eiendeler, Opptjent egenkapital/eiendeler, and Annen Egenkapital.</p>

A6: Datamaskinspesifikasjoner og Python-bibliotek

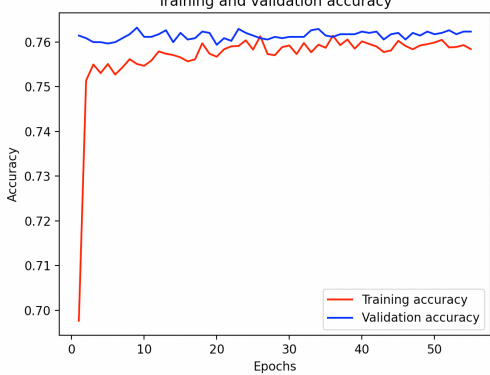
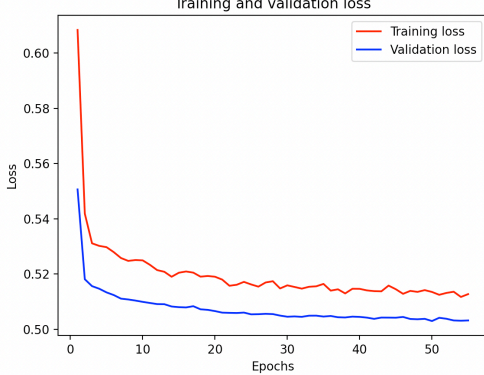
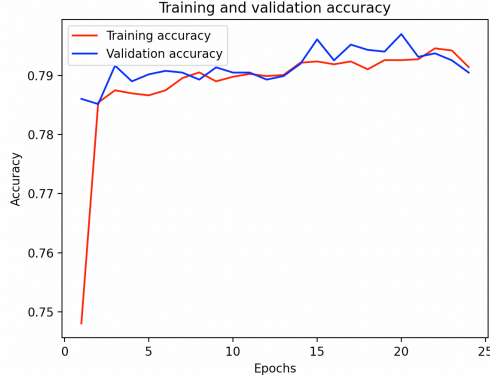
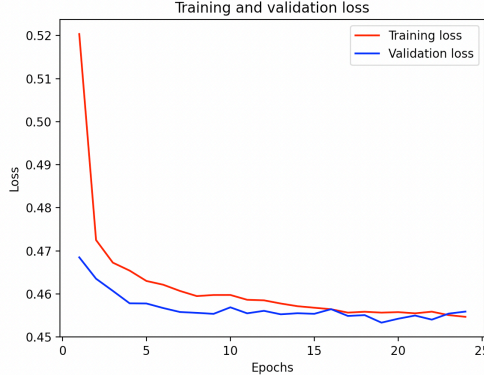
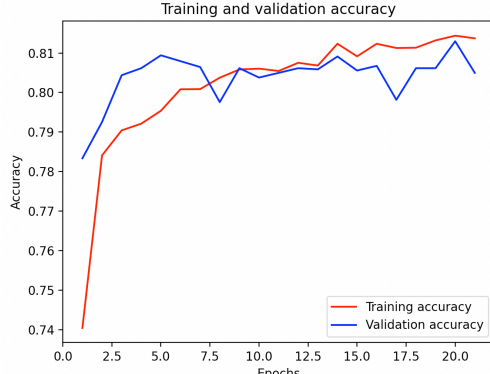
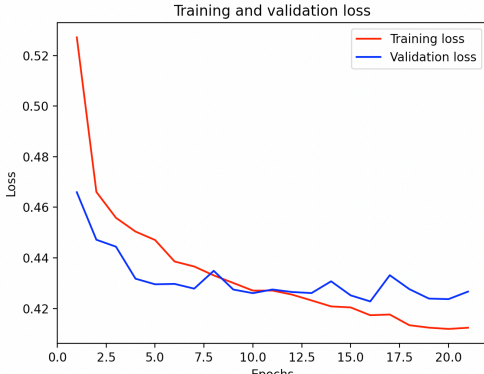
Analysene i oppgaven er utført med en Macbook Air M1 (2020) med følgende spesifikasjoner:

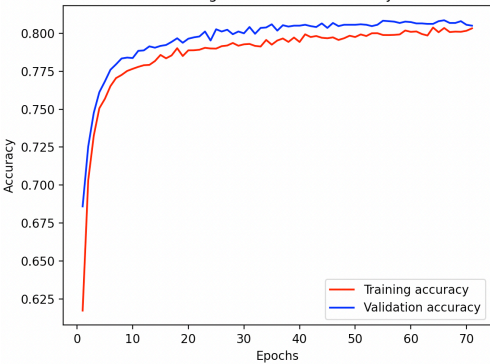
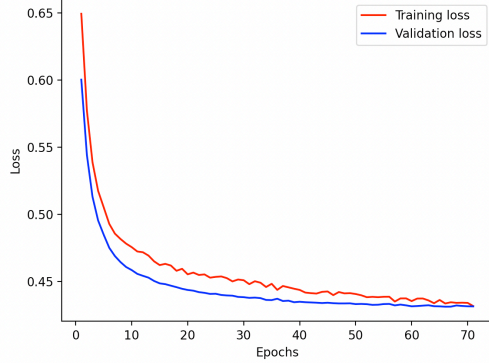
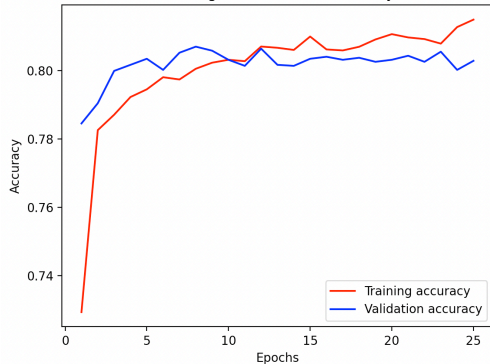
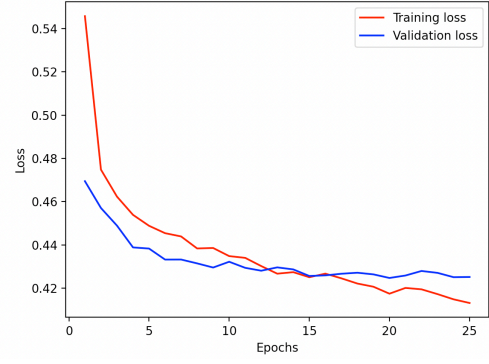
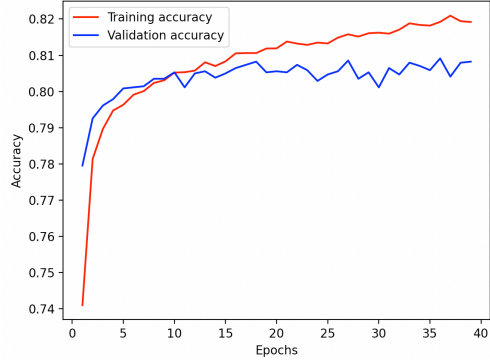
Operativsystem: MacOS Monterey 12.1 – 12.3.1
Prosesor: Apple M1
RAM: 8 GB

Analysene er gjennomført i Visual Studio Code med Python som programmeringsspråk. Tabellen viser versjonen av det benyttede utviklingsverktøyet og Python-versjon med bibliotek.

Programvare	Versjon
Visual Studio Code	1.64 – 1.67
Python	3.9.7
ax-plattform	0.2.4
botorch	0.6.2
keras	2.7.0
matplotlib	3.5.0
numpy	1.21.5
pandas	1.4.0
scikit-learn	1.0.1
shap	0.40.0
tensorflow-macos	2.7.0

A7: Treffprosent og tap på trenings- og valideringssett for nevrale nettverk

Datasett	Treffprosent	Tap
Altman	 <p>Training and validation accuracy for Altman dataset. Training accuracy (red line) starts at 0.70 and rises to approximately 0.76. Validation accuracy (blue line) starts at 0.76 and remains stable around 0.76.</p>	 <p>Training and validation loss for Altman dataset. Training loss (red line) starts at approximately 0.61 and decreases to about 0.51. Validation loss (blue line) starts at approximately 0.55 and decreases to about 0.50.</p>
SEBRA	 <p>Training and validation accuracy for SEBRA dataset. Training accuracy (red line) starts at approximately 0.75 and rises to about 0.79. Validation accuracy (blue line) starts at approximately 0.75 and rises to about 0.79.</p>	 <p>Training and validation loss for SEBRA dataset. Training loss (red line) starts at approximately 0.52 and decreases to about 0.45. Validation loss (blue line) starts at approximately 0.47 and decreases to about 0.45.</p>
Regnskap	 <p>Training and validation accuracy for Regnskap dataset. Training accuracy (red line) starts at approximately 0.74 and rises to about 0.81. Validation accuracy (blue line) starts at approximately 0.78 and rises to about 0.81.</p>	 <p>Training and validation loss for Regnskap dataset. Training loss (red line) starts at approximately 0.53 and decreases to about 0.41. Validation loss (blue line) starts at approximately 0.47 and decreases to about 0.42.</p>

Datasett	Treffprosent	Tap
Regnskap-	<p>Training and validation accuracy</p> 	<p>Training and validation loss</p> 
Regnskap+ Al tman	<p>Training and validation accuracy</p> 	<p>Training and validation loss</p> 
Regnskap+ Al tman-	<p>Training and validation accuracy</p> 	<p>Training and validation loss</p> 