

Anette Håkenstad
Martinius Johansen

Konkursprediksjon av norske SMB

En studie av hvordan Explainable Artificial Intelligence (xAI) øker tolkbarheten til komplekse maskinlæringsmodeller

Masteroppgave i Økonomi og administrasjon
Veileder: Ranik Raaen Wahlstrøm
Mai 2023

Anette Håkenstad
Martinius Johansen

Konkursprediksjon av norske SMB

En studie av hvordan Explainable Artificial Intelligence (xAI) øker tolkbarheten til komplekse maskinlæringsmodeller

Masteroppgave i Økonomi og administrasjon
Veileder: Ranik Raaen Wahlstrøm
Mai 2023

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
NTNU Handelshøyskolen



Kunnskap for en bedre verden

Forord

Det er med stor glede og lettelse at vi presenterer denne masteroppgaven. Masteroppgaven markerer avslutningen av på en utfordrende, men givende reise som studenter på masterprogrammet Økonomi og administrasjon med hovedprofil Business Analytics ved Norges teknisk-naturvitenskapelige universitet (NTNU). Studiet har vært spennende, lærerikt og sosialt, men også til tider krevende. Vi håper og tror at vi nå har tilegnet oss god og nyttig kunnskap som vi kan ta med oss ut i arbeidslivet.

Først og fremst ønsker vi å rette en stor takk til vår veileder, Ranik Raaen Wahlstrøm, for hans uvurderlige støtte, veiledning og engasjement gjennom hele prosessen. Uten hans raske tilbakemeldinger og gode synspunkter i diskusjoner ville oppgaven vært en helt annen.

En spesiell takk går til venner og familie, som har vært der for oss gjennom hele studietiden. Dere har støttet og oppmuntret oss gjennom både oppturer og nedturer, som har vært av stor betydning for oss.

Innholdet i denne oppgaven står for forfatternes regning.

Sammendrag

Å kunne predikere konkurs er hensiktsmessig for flere aktører i finanssektoren, og konkursprediksjon har over mange tiår vært et sentralt emne å undersøke blant akademikere, finansinstitusjoner og bedrifter. I senere år har det blitt utviklet nye og mer komplekse maskinlæringsmetoder som har vist seg å ha svært god prediksjonsevne. På en annen side kjennetegnes disse komplekse maskinlæringsmetodene med lav iboende tolkbarhet, og presenteres i litteraturen som The Black Box Problem (svarte bokser). Som et resultat av denne utfordringen, brukes de komplekse metodene sjelden i praksis, hvor det er de enklere maskinlæringsmodellene med høy iboende tolkbarhet som foretrekkes.

I denne masteroppgaven ser vi derfor nærmere på problemstillingen «Hvordan kan xAI øke tolkbarheten av komplekse maskinlæringsmodeller for konkursprediksjon, og hvor aktuelt er det å ta i bruk dette i finanssektoren?». Dette gjør vi ved å ta i bruk maskinlæringsmetoden XGBoost, som er en trebasert ensemble metode, for konkursprediksjon. For å få et godt bilde på fordelene ved å bruke XGBoost, så velger vi å sammenligne modellenes ytelse med logistisk regresjon, som er en enklere maskinlæringsmetode med god iboende tolkbarhet. Våre modeller trenes og testes på et datasett som representerer norske små- og mellomstore bedrifter i perioden 2006-2019. XGBoost-modellene viser seg i våre resultater å gjøre det bedre enn de logistiske regresjonsmodellene i samtlige perioder med en gjennomsnittlig AUC lik 90% out-of-sample, mot logistiske regresjonsmodeller med en gjennomsnittlig AUC lik 87%. XGBoost har også gjennomgående lavere Brier Score enn logistisk regresjon. XGBoost-modellene har derfor bedre prediksjonsevne enn de logistiske regresjonsmodellene på våre data.

Videre forsøker vi å øke XGBoost-modellenes tolkbarhet ved å anvende Shapley Additive Explanations-rammeverket (SHAP), som er en form for Explainable AI. Først tar vi i bruk SHAP for å gjøre globale forklaringer, hvor vi finner at modellenes generelle logikk er i tråd med økonomisk teori som øker modellenes pålitelighet. Deretter tar vi i bruk SHAP for lokale forklaringer, hvor SHAP gjør det mulig å tolke hvilke variabler som påvirker en enkeltbeslutning og hvilken effekt variablene har på prediksjonen. Totalt sett finner vi at SHAP gjør det mulig tolke både på globalt og lokalt nivå, samtidig som modellens prediksjonsevne opprettholdes. Derfor anser vi SHAP-rammeverket som en effektiv løsning på utfordringen knyttet til svarte bokser, og dermed å gjøre komplekse maskinlæringsmodeller mer anvendbare for konkursprediksjon i finanssektoren.

Abstract

Being able to predict bankruptcy is relevant for multiple actors in the financial sector, and bankruptcy prediction has been a central topic of investigation among academics, financial institutions, and businesses for many decades. In recent years, new and more complex machine learning methods have been developed, showing excellent predictive capabilities. However, these complex machine learning methods are characterized by low inherent interpretability and are referred to as “The Black Box Problem” in the literature. As a result of this challenge, these complex methods are rarely used in practice, where simpler machine learning models with high inherent interpretability are preferred.

In this master’s thesis, we therefore examine the research question «How can xAI enhance the interpretability of complex machine learning models for bankruptcy prediction, and how relevant is it to adopt this in the financial sector?». We accomplish this by using the machine learning method XGBoost, which is a tree-based ensemble method, for bankruptcy prediction. To gain a good understanding of the advantages of using XGBoost, we choose to compare the performance of the models with logistic regression, which is a simpler machine learning method with good inherent interpretability. Our models are trained and tested on a dataset representing Norwegian small and medium-sized enterprises from 2006 to 2019. In our results, XGBoost models outperform logistic regression models in all periods, with an average out-of-sample AUC of 90%, compared to logistic regression models with an average AUC of 87%. XGBoost also consistently exhibits lower Brier Score than logistic regression. Therefore, XGBoost models demonstrate better predictive capabilities than logistic regression models on our data.

Furthermore, we attempt to enhance the interpretability of XGBoost models by applying the Shapley Additive Explanations framework (SHAP), which is a form of Explainable AI. First, we use SHAP to provide global explanations, finding that the models’ general logic aligns with economic theory, thus increasing their reliability. Next, we employ SHAP for local explanations, allowing us to interpret which variables influence an individual decision and the effects these variables have on the prediction. Overall, we find that SHAP enables interpretability at both global and local levels while maintaining the models’ predictive performance. Therefore, we consider the SHAP framework an effective solution to the challenge of black boxes and making complex machine learning models more applicable for bankruptcy prediction in the financial sector.

Innhold

1	Introduksjon	1
1.1	Tema og aktualisering	1
1.2	Forskningsspørsmål	4
1.3	Oppgavens struktur	5
2	Litteraturgjennomgang og teoretisk bakgrunn	6
2.1	Bruk av kunstig intelligens i finanssektoren	6
2.1.1	Tradeoff mellom tolkbarhet og nøyaktighet	6
2.1.2	KI i finanssektoren i dag	8
2.2	Konkursprediksjon	9
2.3	Trebaserte modeller	11
2.4	The Black Box Problem og Explainable AI	13
2.5	Tolkning	14
2.5.1	Global forklaring	15
2.5.2	Lokal forklaring	15
2.5.3	Tolkbarhetsmetoder	16
2.6	Variabler	17
2.6.1	Bruk av regnskapsbaserte variabler	17
2.6.2	Metoder for variabelseleksjon	19
3	Datasett	21
3.1	Konkurs som responsvariabel	21
3.2	Variabler	21
3.3	Filtrering av data	22

3.4	Winzorizing	23
3.5	Behandlet datasett	23
4	Metode	25
4.1	Treningssett og testsett	25
4.2	Håndtering av ubalanserte data	26
4.3	Estimeringsmetoder	27
4.3.1	Logistisk regresjon	27
4.3.2	Extreme Gradient Boosting	28
4.4	Tolkning av modeller	30
4.4.1	LASSO	31
4.4.2	SHAP	32
4.5	Evaluering av modell	33
4.5.1	AUC	34
4.5.2	Brier score	34
5	Resultat og diskusjon	36
5.1	Variabelseleksjon ved LASSO	36
5.2	Modellestimering	38
5.2.1	Logistisk regresjon	39
5.2.2	XGBoost	39
5.2.3	Sammenligning av logistisk regresjon og XGBoost	40
5.3	SHAP	41
5.3.1	Global forklaring	41
5.3.2	Lokal forklaring	46
5.4	Anvendelse i praksis	53

6 Konklusjon	56
6.1 Svakheter ved studien	57
6.2 Videre forskning	58
Bibliografi	60
Appendix	73
A Regnskapsposter i opprinnelig datasett	73
B Variabler i endelig datasett	77
C Forward selection	79
D LASSO stiplott	80
E SHAP beeswarm for hvert testår	90

Figurer

1	Utvikling i lånegjelden til ikke-finansielle institusjoner hos banker og antall konkurser i Norge pr. år	1
2	Hvordan ulike maskinlæringsmetoder plasserer seg i forhold til prediksjonsevne og tolkbarhet	3
3	Lokal forklaring. Konkursbedrift med lav sannsynlighet for konkurs. .	47
4	Lokal forklaring. Konkursbedrift med høy sannsynlighet for konkurs. .	48
5	Lokal forklaring. Ikke-konkursbedrift med lav sannsynlighet for konkurs.	50
6	Lokal forklaring. Ikke-konkursbedrift med høy sannsynlighet for konkurs.	50
7	Stiplott for testår 2010	80
8	Stiplott for testår 2011	81
9	Stiplott for testår 2012	82
10	Stiplott for testår 2013	83
11	Stiplott for testår 2014	84
12	Stiplott for testår 2015	85
13	Stiplott for testår 2016	86
14	Stiplott for testår 2017	87
15	Stiplott for testår 2018	88
16	Stiplott for testår 2019	89
17	SHAP beeswarm plott for testår 2010	90
18	SHAP beeswarm plott for testår 2011	90
19	SHAP beeswarm plott for testår 2012	91
20	SHAP beeswarm plott for testår 2013	91
21	SHAP beeswarm plott for testår 2014	92

22	SHAP beeswarm plott for testår 2015	92
23	SHAP beeswarm plott for testår 2016	93
24	SHAP beeswarm plott for testår 2017	93
25	SHAP beeswarm plott for testår 2018	94
26	SHAP beeswarm plott for testår 2019	94

Tabeller

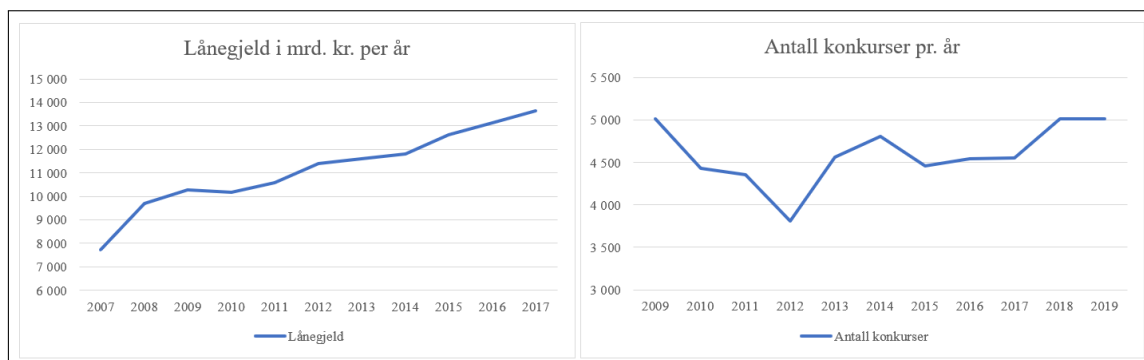
1	Deskriptiv statistikk for behandlet datasett	24
2	Tuning av hyperparamtere for XGBoost	30
3	Regresjonskoeffisienter og tilhørende z-verdier for hvert testår	37
4	Modellenes prediksjonsevne i AUC og Brier-verdier	39
5	Standardiserte SHAP-verdier	42
6	Forventet verdi for variablene i figur 4.	48
7	Forventet verdi for variablene i figur 6.	51
8	Regnskapsposter i det opprinnelige datasettet	73
9	Liste over variabler som er benyttet i modellene	77
10	Rolling window	79

1 Introduksjon

1.1 Tema og aktualisering

Når en bedrift går konkurs har dette store konsekvenser for alle involverte parter. Kreditorer påvirkes i stor grad ved at bedrifter har utestående gjeld de ikke evner å betale, og de risikerer derfor å gå med tap. Til tross for at kreditorer i dag som oftest anvender ulike modeller for sannsynlighetsberegninger for mislighold av forpliktelse i sin utlånsstrategi, vil det alltid være til stede en risiko for uventede tap som er umulig å beregne på forhånd. Slike uventede tap kan være grunnlag for at store verdier går tapt for kreditorerne. Ved begjæring av konkurs vil også avtaler og kontrakter knyttet til virksomheten termineres, og dermed føre til økonomiske og operasjonelle ringvirkninger for alle involverte. Dette gjelder leverandører, kunder og samarbeidspartnere som rammes av økonomiske tap når bedriften ikke evner å betale sine utestående forpliktelser, samt tap av viktige samarbeidsavtaler. Videre vil ansatte miste kontraktene sine ved konkurs, som er særlig problematisk for hver enkelt ansatt i et presset arbeidsmarked.

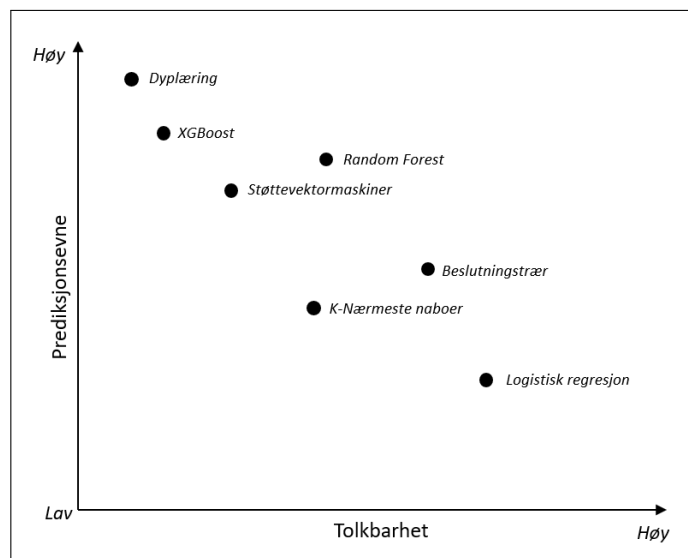
Å kunne forutsi konkurs er av vesentlig interesse for en rekke aktører i finanssektoren, i form av investorer, kreditorer, finansmarkedet og samfunnet som helhet siden konkurser fører til tap for alle parter i større eller mindre grad (Alaka mfl., 2017; Van Gestel mfl., 2003). Grafen til venstre i figur 1 viser at ikke-finansielle selskapers lånegjeld hos bankene har økt betydelig fra 2007 til 2017. Dessuten kan man se at antall bedrifter som går konkurs per år holder seg relativt stabilt. Jo større lån bankene låner ut, jo større blir deres tapspotensial.



Figur 1: Utvikling i lånegjelden til ikke-finansielle institusjoner hos banker og antall konkurser i Norge pr. år (Statistisk-sentralbyrå, 2022a,b).

Konkurs er med andre ord et fenomen som gir alvorlige konsekvenser for en rekke aktører innenfor finanssektoren. Det er derfor, til tross for at konkursprediksjon er et velstudert fagområde, fortsatt viktig med utvikling og nye funn i litteraturen. I Norge blir kunstig intelligens (KI) brukt i økende grad (Lervik, 2022), hvor enklere maskinlæringsmetoder og andre statistiske metoder er mest brukt. Dette ser vi blant annet i SEBRA-modellen som Norges Bank og Finanstilsynet anvender for å predikere sannsynlighet for konkurs og mislighold for norske bedrifter. SEBRA-modellen tar i bruk en generalisert logistisk regresjonsmodell med god iboende tolkbarhet, og relativt god prediksjonsevne (Bernhardsen og Larsen, 2007).

Komplekse maskinlæringsmodeller er i svært liten grad brukt for å predikere konkurs og annen mislighold i dag, som et resultat av at disse metodene ofte er vanskelige å tolke (Lervik, 2022). Dette er til tross for at komplekse modeller ofte har bedre prediksjonsevne, og at en marginal økning i prediksjonsevnen kan gi signifikant økning i bankers lønnsomhet (Trinkle og Baldwin, 2007). Vanligvis omtales tolkbarhetsutfordringene knyttet til de komplekse modellene som «The Black Box Problem». Dette refererer til manglende transparens knyttet til beslutningene som gjøres av de komplekse modellene. Dette gjør det utfordrende å ta i bruk disse metodene innenfor finanssektoren, da sektoren er sterkt regulert av blant annet General Data Protection Regulation (GDPR) som krever transparens og logiske forklaringer knyttet til automatiserte beslutninger (EUR-Lex, 2016). Det er likevel viktig å nevne at GDPR kun gjelder privatpersoner og ikke hele bedrifter, men at reguleringen gir et bilde på viktigheten av transparens og tolkbarhet, og hvorfor enklere modeller ofte blir tatt i bruk i praksis.



Figur 2: Her vises hvordan ulike maskinlæringsmetoder plasserer seg i forhold til prediksjonsevne og tolkbarhet (Arrieta mfl., 2020; Ciatto mfl., 2020)

I figur 2 presenteres hvordan ulike maskinlæringsmetoder innenfor maskinlæring plasserer seg i forhold til tolkbarhet og prediksjonsevne. Her ser vi at dyplæring, derunder nevrale nettverk, og Extreme Gradient Boosting (XGBoost) plasserer seg høyt på prediksjonsevne, men ligger relativt lavt på tolkbarhet. På andre siden av spekteret ligger de enkle modellene med høyere tolkbarhet og lavere prediksjonsevne, herunder blant annet logistisk regresjon.

Som en løsning på tolkbarhetsutfordringene knyttet til bruk av komplekse maskinlæringsmetoder, presenteres Explainable AI (xAI) (Castelvecchi, 2016). xAI sikter på å øke tolkbarheten til maskinlæringsmodeller, samtidig som det er rettet fokus på å opprettholde modellens prediksjonsevne. En sentral teknikk innenfor xAI er SHapley Additive exPlanations (SHAP), som blir presentert av Lundberg og Lee (2017). Denne teknikken gjør det mulig å undersøke variabelviktighet for modeller gjennom global forklaring, så vel som å tolke hvordan enkeltbeslutninger er gjennomført i modellen ved hjelp av SHAP-verdier. I denne studien tar vi i bruk SHAP som et verktøy for å øke tolkbarheten til XGBoost-modeller.

Per 2023 utgjør små og mellomstore bedrifter (SMB) i overkant av 99% av alle private virksomheter på tvers av alle næringer i Norge (Statistisk-sentralbyrå, 2023). I 2019 hadde 95% av alle bedrifter mindre enn 50 ansatte (Næringsdepartementet, 2019). Videre viste tall i en rapport fra DNB og NyAnalyse i 2018 at SMB sto for en verdiskapning på 700 milliarder kroner, som representerer 42% av landets verdiskapning, samt 62% av norsk næringslivs sysselsetting (Sivam mfl., 2018). Til

tross for at SMB utgjør en så stor andel av alle virksomheter, er det gjort få studier i forbindelse med konkursprediksjon av disse. De fleste studier gjøres på store og børsnoterte bedrifter, da det er mer tilgjengelig informasjon om disse bedriftene enn hva som er tilfelle for SMB (Paraschiv mfl., 2021). Det vil være fordelaktig både for banker og SMB å trene opp gode og nøyaktige konkursprediksjonsmodeller, da dette vil kunne bidra til å redusere bankenes tapsrisiko, samtidig som det gir bedre muligheter for finansiering for SMB med gode lånebetingelser (Tobback mfl., 2017).

1.2 Forskningsspørsmål

På bakgrunn av at norske bedrifters gjeld øker og konsekvensene som følger av konkurs, anser vi det som viktig å kunne forutsi hvilke bedrifter som kommer til å gå konkurs i fremtiden. I følge tidligere litteratur viser det seg at mer komplekse modeller øker prediksjonsevnen, men at dette går på bekostning av tolkbarheten. Målet med denne studien er derfor å undersøke om hvorvidt SHAP-rammeverket kan brukes til å tolke konkursprediksjonene som komplekse maskinlæringsmodeller gjør. Vi har dermed følgende problemstilling:

Hvordan kan xAI øke tolkbarheten av komplekse maskinlæringsmodeller for konkursprediksjon, og hvor aktuelt er det å ta i bruk dette i finanssektoren?

For å undersøke problemstillingen predikerer vi først konkurs ved bruk av logistisk regresjon, som er den vanligste metoden for konkursprediksjon, hvor vi gjennomfører variabelseleksjon ved bruk av Least Absolute Shrinkage and Selection Operator (LASSO). Videre sammenligner vi denne med den mer komplekse maskinlæringsmetoden XGBoost, for å etablere at sistnevnte predikerer bedre, i tråd med presentasjonen i figur 2. Dermed viser vi hvordan en modell med høy prediksjonsevne og lav iboende tolkbarhet allikevel kan tolkes med xAI. Vi undersøker deretter om hvorvidt SHAP-rammeverket kan brukes for tolkning av komplekse maskinlæringsmodeller i finanssektoren. Vi velger å undersøke XGBoost i stedet for dyplæring, som i figur 2 presenteres som den metoden med høyest prediksjonsevne. Dette valget er i stor grad basert på det faktum at prediksjon med dyplæring i kombinasjon med SHAP krever svært mye datakraft, som anses som lite hensiktsmessig innenfor tidsrammen og ressursene vi har tilgjengelig for denne studien.

Med denne studien gir vi flere bidrag til litteraturen og praktikere. For det første gir vi en presentasjon av hvordan anvendelse av SHAP-rammeverket som tolkningsme-

tode kan øke mer komplekse maskinlæringsmetoders tolkbarhet, og hvorvidt dette kan tas i bruk i praksis i finanssektoren. For det andre bidrar studien med å bekrefte at mer komplekse maskinlæringsmodeller generelt sett gir bedre prediksjonsevne enn enklere modeller. For det tredje bidrar studien med økt litteratur innen konkursprediksjon som fagområde, også i Norge. Som tidligere nevnt i kapittel 1.1 er det gjort flest studier på store og børsnoterte bedrifter innenfor konkursprediksjon. Vi bidrar derfor også med funn knyttet til SMB som det er begrenset litteratur på. Uansett er hovedbidraget i studien presentasjonen og diskusjonen rundt hvordan finanssektoren kan ta i bruk SHAP-rammeverket som forklaring på de svarte boksene som følger med komplekse maskinlæringsmodeller.

1.3 Oppgavens struktur

I kapittel 2 presenterer vi studiens teoretiske bakgrunn og tidligere litteratur knyttet til konkursprediksjon. Først presenteres bruk av kunstig intelligens i finanssektoren i dag, før vi ser på tidligere studier av konkursprediksjon. Videre går vi igjennom trebaserte modeller og tolkningen av mer komplekse maskinlæringsmetoder. Til slutt presenteres variabelenes sentrale rolle for konkursprediksjon.

Kapittel 3 gir en detaljert beskrivelse av datasettet som brukes i studien. Først presenteres dataene og definisjonen av konkurs. Videre gis en nærmere forklaring på datasettets variabler og hvordan datasettet er behandlet. Til slutt presenteres det behandlede datasettet.

I kapittel 4 forklares først oppdelingen av treningssett og testsett og håndteringen av ubalansert data. Videre presenteres maskinlæringsmetodene logistisk regresjon og XGBoost, samt hvordan disse tolkes. Til slutt utleder vi hvordan vi evaluerer modellene med AUC og Brier score.

Kapittel 5 presenterer resultatene av studien som legger grunnlag for diskusjon. Først presenteres resultatene og en drøfting av variabelseleksjon for logistisk regresjon. Deretter presenteres modellenes prediksjonsevne og en sammenligning av maskinlæringsmetodene. Til slutt presenteres og diskuteres resultatet av SHAP knyttet til global og lokal forklaring.

Til slutt i kapittel 6 legges det frem en konklusjon basert på våre funn, før vi ser på svakheter ved studien. I siste delkapittel presenteres forslag til videre forskning.

2 Litteraturgjennomgang og teoretisk bakgrunn

I dette kapittelet skal vi i hovedsak gå igjennom tidligere litteratur som er gjort innenfor konkursprediksjon, samt andre studier som er relevant for denne studien.

2.1 Bruk av kunstig intelligens i finanssektoren

Komplekse maskinlæringsmodeller har vist seg som svært effektive til å løse ulike problemstillinger for store datamengder innenfor datavitenskap og kunstig intelligens, og anses å potensielt kunne bidra til å øke bedrifters lønnsomhet (Brynjolfsson mfl., 2011). I tråd med den økte digitaliseringen i næringslivet og samfunnet forøvrig, brukes kunstig intelligens (KI) i større grad som et sentralt element i bedrifters overordnede forretningsstrategi med mål om økt verdi og lønnsomhet (Arrieta mfl., 2020). Derfor anses ofte finansielle prognoseverktøy som eksempelvis prediksjonsmodeller for konkurs som en verdifull ressurs for finansinstitusjoner, bedrifter og dets interessenter.

2.1.1 Tradeoff mellom tolkbarhet og nøyaktighet

Son mfl. (2019) presenterer to kriterier for at en modell for konkursprediksjon skal anses som effektiv: (1) modellen må være nøyaktig i sine prediksjoner, og (2) modellen må være tolkbar. På en annen side er det ofte en avveining mellom hvor godt en modell kan predikere og hvor enkelt det er å tolke modellens prediksjoner (Dovsilovic mfl., 2018). Denne avveiningen er ofte basert på graden av kompleksitet i modellen. En mer kompleks modell som gir bedre prediksjonsevne kan være vanskelig å tolke og forklare, i forhold til en enklere modell med dårligere prediksjonsevne. Det er derfor viktig å finne en balanse mellom tolkning og prediksjonsevne, avhengig av formålet med modellen og bruksområde (Wanner mfl., 2021). I visse situasjoner vil tolkning være viktigere enn prediksjonsevne, mens i andre situasjoner kan prediksjonsevnen være av større betydning. I for eksempel helsevesen, forsvar og finansbransjen er tolkbarhet en viktig faktor for å kunne bygge tillit til modellen og dens resultater, da bakgrunn for beslutninger er sentralt og påvirker både privatpersoner og organisasjoner. Om man har en modell med spesielt nøyaktige prediksjoner, men det er vanskelig å tolke hvordan modellen har kommet til sine resultater, kan det være utfordrende for fagpersoner å ha tillit til modellen og foreta beslutninger på bakgrunn av resultatene. Generelt tyder det på at det ofte prefereres enklere maskinlæringsal-

goritmer som logistisk regresjon med høy grad av iboende tolkbarhet (Rudin, 2019). Dette er til tross for at man som oftest oppnår høyere prediksjonsevne ved bruk av mer avanserte maskinlæringsmetoder (Wanner mfl., 2020).

Det er likevel viktig å være klar over at en enkel modell med god iboende tolkbarhet ikke nødvendigvis er mindre nøyaktig enn en mer kompleks modell (James mfl., 2013). I visse tilfeller kan faktisk enkle modeller gi tilnærmet like gode resultater som en mer kompleks modell, og dermed være å foretrekke. Dette utartes gjerne i situasjoner hvor man har lite data tilgjengelig. På den andre siden ses gjerne tendensen til at komplekse modeller har høyere prediksjonsevne på store og mer komplekse datasett, da disse modellene har komplekse sammenhenger mellom variabler og indre funksjoner (Wanner mfl., 2021).

Det er med andre ord ingen «one-size-fits-all»-løsning innenfor denne tematikken. Innenfor finanssektoren er det ulike anvendelser av finansielle prognoseverktøy, hvor det også er ulikt behov for nøyaktighet og tolkbarhet. For banker som låner ut penger vil det være størst fokus på nøyaktighet, da de risikerer å tape penger ved en eventuell konkurs eller mislighold av gjeld. Samtidig er det mindre relevant å se hvordan modellen har foretatt enkeltbeslutninger. For ulike bedrifter er det derimot av interesse å se hvilke faktorer som påvirker prediksjonen, for å finne hvilke beslutninger som bør tas for å redusere konkurssannsynligheten. Bedriftene har dermed et større fokus på tolkbarheten av prediksjonen. Bracke mfl. (2019) formulerer tre sentrale spørsmål for bedrifters interesser, knyttet til tolkning av prediksjonsmodeller:

- Hva er det som generelt driver forklaringene?
- Hvilke variabler, med hvilken effekt, er av betydning for de individuelle prediksjonene?
- Hvordan fungerer modellen, og kan dette enkelt forklares?

Paraschiv mfl. (2021) presenterer to typer kostnader som konsekvens av feilprediksjon av konkurs: (1) predikerer at en bedrift ikke går konkurs selv om den faktisk går konkurs, og (2) predikerer at en bedrift går konkurs selv om den faktisk ikke går konkurs. For førstnevnte vil kostnaden ved feilprediksjon være at bankene låner ut penger til bedrifter som går konkurs og misligholder gjelden, og dermed taper faktiske penger som er lånt ut. Den andre typen kostnad dreier seg i større grad om potensielle inntekter som går tapt, som resultat av at potensielle bankkunder velger å gå til andre banker. Flere studier finner at det er kostnaden knyttet til den første

feilprediksjonen som oftest har størst negativ konsekvens for bankene (Stein, 2005; Balcaen og Ooghe, 2006; Agarwal og Taffler, 2007). En prediksjonsmodell kan ikke eliminere begge disse typene feilprediksjon, og derfor vil en liten økning i prediksjonsnøyaktigheten, i f.eks. evalueringskriteriet AUC, kunne ha stor påvirkning på bankens lønnsomhet (Trinkle og Baldwin, 2007).

2.1.2 KI i finanssektoren i dag

Lervik (2022) forteller i sin studie at det har vært en sterk økning i satsing på kunstig intelligens (KI) i den norske finanssektoren de siste 3-5 årene, og at den norske finanssektoren leder an med høy bruk av KI-teknologi sammenlignet med europeisk finansnæring. Han finner videre at de fleste aspekter av banksektoren har tatt i bruk kunstig intelligens teknologi i en eller annen form, hvor komplekse modeller brukes i mindre grad. Bankene forklarer at hovedgrunnene til dette er knyttet til utfordringene til modellenes tolkbarhet, samt sikkerheten knyttet til rettferdig behandling av kunder. Det nevnes derfor at anvendelse av xAI anses som avgjørende for mer utstrakt bruk av KI i den norske finanssektoren. Kompetansen knyttet til bruk og vedlikehold av KI-systemer nevnes også som en hindring for bruk av mer komplekse former av KI i finanssektoren, samt at det er arbeidskrevende å bygge opp en nødvendig teknisk infrastruktur. Videre legges det også vekt på mangelen på data av god kvalitet, samt forståelsen for hvordan store mengder data kan brukes og utnyttes på en god og hensiktsmessig måte (Lervik, 2022).

Finanssektoren er videre svært regulert av lover og reguleringer som setter krav om høy transparens og ansvarlighet. Det er svært viktig at beslutninger som tas kan forklares enkelt, og at man har innsikt i hvordan og på hvilket grunnlag en beslutning er tatt (EUR-Lex, 2016). Komplekse maskinlæringsmodeller er derimot som nevnt vanskelige å tolke og forstå, og kan være utsatt for feil som er vanskelig å oppdage og justere. En annen konkret regulering som krever transparens og åpenhet i beslutningsprosesser er det internasjonale regulatoriske rammeverket Basel 3. Dette rammeverket er utarbeidet for å øke stabiliteten og robustheten i finanssektoren gjennom fastsettelse av krav til banker om blant annet kapital, likviditet og gjeldsgrad (BIS, 2011). En av bestemmelsene i rammeverket er at det skal gjennomføres omfattende risikovurdering, og at resultatene av risikovurderingen skal dokumenteres og rapporteres. I tillegg krever avtalen at banker må ha tilstrekkelig kvalitet og mengde av data for å kunne gjennomføre gode risikovurderinger og beslutninger. Dette krever god datakvalitet, som kan dokumenteres og verifiseres. Avtalen krever

med andre ord større åpenhet og transparens i beslutningstaking og risikovurdering, som i mindre grad gjør det mulig å ta i bruk komplekse og mindre tolkbare maskinlæringsmodeller.

Videre er personvernreguleringen gjennom GDPR sentral. Kapittel 3 i GDPR regulerer beslutninger som påvirker privatpersoner, hvorav spesielt artikkel 22 er sentral. Artikkelen gir individet rett til ikke å bli gjenstand for beslutninger som er basert på automatisk behandling. Videre utledes det i artikkelen at individet har rett på en forklaring på logikken bak en enkeltbeslutning, og hvilken betydning og konsekvenser dette har for vedkommende (EUR-Lex, 2016).

Som nevnt er det en utfordring i finanssektoren at dataene som brukes ofte er komplekse og ustrukturerte. Dette bunner i at innsamling og analyser av større mengder data fra ulike kilder ofte kommer i ulike format, og at en standardisering av dataene ofte krever avansert datateknologi og kompetanse på området. I tillegg anses økonomi som et komplekst fagområde med en rekke faktorer som påvirker en rekke ulike variabler (Arthur, 2009). Det er derfor utfordringer knyttet til at det er vanskelig å fange opp og identifisere disse i økonomiske prediksjonsmodeller, hvorav politiske og makroøkonomiske forhold (Stock og Watson, 1992). Det er med andre ord vanskelig å utvikle modeller som klarer å predikere fremtidige hendelser på en nøyaktig og realistisk måte. Maskinlæringsmodellene trenger store mengder data for å fungere effektivt, og finanssektoren har ofte ikke tilstrekkelig data til å trene og teste mer komplekse modeller (Lervik, 2022).

2.2 Konkursprediksjon

Konkursprediksjon har vært et aktivt forskningsområde i flere tiår. I begynnelsen fokuserte en rekke forskere hovedsakelig på å undersøke finansielle indikatorer, som inntjening, likviditet og gjeld, for å predikere konkurs (Beaver, 1966; Altman, 1968; Ohlson, 1980; Zmijewski, 1984). Allerede på 1960-tallet forsøkte Beaver (1966) å identifisere faktorer som kan brukes for å predikere konkurs med en univariat modell. De første modellene som kom til bruker som oftest diskriminantanalyse. Dette ser vi blant annet i Altman (1968), som presenterte en multipel diskriminant analyse, for å unngå inkonsistensutfordringene som følger med en univariat modell. I etterkant av denne studien ble multipel diskriminantanalyse (MDA) hyppig brukt i flere studier for konkursprediksjon, hvorav blant annet Deakin (1972), Edmister (1972) og Altman mfl. (1977). Denne metoden var i mange år den mest brukte metoden

innenfor konkursprediksjon (Altman og Sabato, 2007).

På en annen side blir det gjennomgående påpekt i flere studier at de grunnleggende antakelsene om normalfordelte variabler og gruppespredningsmatriser i MDA ofte brytes ved konkursprediksjon (Altman og Sabato, 2007). Logistisk regresjon ble imidlertid først tatt i bruk innenfor konkursprediksjon av Ohlson (1980), og har over tid vært den mest foretrukne metoden for konkursprediksjon (Shi og Li, 2019). Populariteten forklares av en kombinasjon av god tolkbarhet, god prediksjonsevne og mindre restriktive antagelser enn MDA (Paraschiv mfl., 2021).

I følge Shi og Li (2019) er de mest brukte maskinlæringsmetodene for konkursprediksjon MDA, logistisk regresjon og nevralt nettverk. MDA og logistisk regresjon er relativt enkle statistiske klassifiseringsmetoder, som samtidig gir rimelig gode konkursprediksjoner (Jones mfl., 2015). Shumway (2001) presenterer at diskrete hazard-modeller, som er en type logistisk regresjon, er mer nøyaktig enn MDA. Denne metoden tar i bruk flere årsregnskaper for hver bedrift i modellen, og ikke bare årsregnskap for foregående år (Shumway, 2001). I nyere tid har både nevralt nettverk og så kalte «new age» statistiske modeller økt i popularitet (Jones mfl., 2017), som et resultat av at disse modellene er mer komplekse og fanger ikke-lineariteter (Kumar og Ravi, 2007). Nevrale nettverk ble tatt i bruk til konkursprediksjon i 1990 med gode resultater (Laitinen og Kankaanpää, 1999). «New age» modellene inkluderer blant annet en rekke ensemble metoder som Random Forest, Gradient Boosting og AdaBoost (Jones mfl., 2017). Jones mfl. (2017) anbefaler heller å ta i bruk disse modellene for konkursprediksjon, fordi de predikerer betydelig bedre enn både MDA og logistisk regresjon. Disse funnene støttes av Alaka mfl. (2017) som finner at både nevralt nettverk og trebaserte ensemble modeller presterer bedre og er mer nøyaktige enn logistisk regresjon. Til tross for at studier viser at disse modellene har bedre prediksjonsevne, blir metodene ikke brukt hos banker da det oppstår problemer knyttet til å tolke modellens fremgangsmåte og resultater, som strider mot personvernsreguleringen GDPR (EUR-Lex, 2016).

I og med at finanssektoren ofte bruker logistisk regresjon til konkursprediksjon blir metoden også hyppig brukt i studier til å sammenlignes opp mot andre estimeringsteknikker (Bernhardsen og Larsen, 2007). Dette kan vi eksempelvis se i Min og Jeong (2009), Zhang mfl. (1999) og Premachandra mfl. (2009). I de fleste studiene ser man at logistisk regresjon gjør gode prediksjoner med relativt god tolkbarhet, men at de mer avanserte maskinlæringsmetodene gjør det jevnt over bedre, dog med mindre tolkbarhet.

Majoriteten av studier på konkursprediksjon er gjennomført på store og børsnoterte bedrifter i USA. I Norge er det derimot gjort relativt få studier på konkursprediksjon. Bernhardsen og Larsen (2007) utvikler i sin studie en prediksjonsmodell for konkurs med data fra SEBRA-databasen ved å bruke en generalisert form for logistisk regresjon. Modellen brukes i dag av den norske sentralbanken, Norges Bank, og finansilsynet for å predikere konkurs. Til forskjell fra majoriteten av tidligere studier på konkursprediksjon, baserer modellen seg på små- og mellomstore bedrifter (Bernhardsen og Larsen, 2007). SEBRA-modellen gir en nøyaktighet på i overkant av 82%. I nyere tid tar Paraschiv mfl. (2021) utgangspunkt i norske SMB, hvor de presenterer konkursprediksjon ved hjelp av feature selection metoder, kombinert med modellestimering ved bruk av logistisk regresjon og nevralt nettverk.

Kolari og Shin (2006) finner at utlån til små bedrifter har en sterk positiv effekt på banklønnsomhet, til tross for at de finner at utlån til SMB er mer risikabelt enn til større bedrifter. Det er derfor hensiktsmessig for banker å utvikle kredittrisikomodeller som er spesifikt rettet mot SMB, med hensikt å minimere fremtidige potensielle tap (Altman og Sabato, 2007). Altman og Sabato (2007) estimerer en kredittrisikomodel spesifikt for SMB, ved hjelp av fem finansielle variabler og logistisk regresjon. I studien finner de at den logistiske regresjonsmodellen har en prediksjonsnøyaktighet på nesten 30% høyere enn en standard MDA-modell (Altman og Sabato, 2007).

2.3 Trebaserte modeller

Trebaserte modeller, eller beslutningstrær, bruker en trestruktur for å gjøre beslutninger for klassifiseringsproblemer (James mfl., 2013). Beslutningstreet bygges opp fra en rot-node, som representerer det opprinnelige problemet som skal løses, og ut i en serie av interne noder som utfører forskjellige tester på input-dataene. Hver intern node representerer en beslutning, hvor det gjennomføres en test på en bestemt egenskap. Ut fra disse testresultatene fortsetter modellen videre til to eller flere viderekoblede noder. Avslutningsvis ender beslutningstreet i en terminalnode, som representerer den endelige klassifiseringsbeslutningen, og dermed løsningen på det opprinnelige problemet (Breiman, 2017).

Beslutningstrær brukes for både regresjon- og klassifiseringsproblemer, og ble først introdusert av Breiman (2017) som CART (klassifikasjons- og regresjonstrær). CART viser seg å være effektive innenfor blant annet konkursprediksjon, særlig når det kom-

mer til identifikasjon av viktige faktorer som påvirker konkurrisikoen (Kumar og Ravi, 2007). Videre er CART-modeller nyttige for utvikling av risikoprofiler knyttet til bedrifter, og dermed predikere konkurrisiko basert på finansielle data (Olson mfl., 2012). Olson mfl. (2012) presenterer en sammenligningsstudie av ulike maskinlæringsmetoder innenfor konkursprediksjon, hvor blant annet beslutningstrær, støttevektormaskiner og nevrale nettverk sammenlignes, hvor beslutningstrær anses som relativt mest nøyaktig på de gjeldende dataene (Olson mfl., 2012).

Breiman (2017) fremhever modellenes tolkbarhet som den største fordelen ved bruk av CART, men presiserer samtidig at dette typisk skjer på bekostning av modellens prediksjonsevne. I utgangspunktet regnes enkle beslutningstrær som relativt mindre kompliserte modeller. Som nevnt i kapittel 2.1.1 tenderer enkle modeller å ha svakere ytelse enn mer kompliserte modeller. Som en løsning på ytelsesproblemet for beslutningstrær, blir ensemble metoder presentert (Jones mfl., 2017). Disse har vist seg spesielt effektive i konkursprediksjon. Ensemble metoder kombinerer flere beslutningstrær for å oppnå bedre ytelse enn ved bruk av et enkelt beslutningstre (Tsai mfl., 2014). Den grunnleggende idéen bak bruk av ensemble metoder er å kombinere styrker og svakheter fra ulike modeller, slik at man reduserer risiko for overtilpasning og oppnår bedre resultater enn de individuelle modellene. Vi skiller generelt mellom tre typer ensemble metoder (James mfl., 2013):

- **Bagging** (Bootstrap aggregating): En teknikk som tar i bruk flere uavhengige modeller som er trent på forskjellige delmengder (subset) av treningssettet. De individuelle modellene kombineres deretter ved bruk av gjennomsnitt, median eller andre aggresjonsfunksjoner.
- **Boosting**: En teknikk som fokuserer på å bruke flere individuelle modeller for å forbedre samlet ytelse. Dette gjøres ved å trene en serie individuelle modeller, hvor hver modell forsøker å korrigere feilene til foregående modell. Hver modell får dermed større vekt på eksempler som tidligere modeller har feilklassifisert. Den endelige prediksjonen er en kombinasjon av alle de individuelle modellene.
- **Random Forest**: En teknikk som bruker en kombinasjon av beslutningstrær som har blitt trent på en tilfeldig delmengde av dataene, og tilfeldig seleksjon av funksjoner. Denne kombinasjonen gjør at Random Forest reduserer risikoen knyttet til overtilpasning, samtidig som modellens ytelse øker.

Boosting er en populær form for ensemble metode som følge av måten den evner å forbedre ytelsen til svake modeller ved å kombinere flere svake modeller til én ster-

kere modell (Dietterich, 2000). En sterkere modell evner i større grad å generalisere problemer. I tillegg reduserer boosting risikoen for overtilpasning ved at de vekter datapunkter som feilklassifiseres for hver iterasjon, og dermed tvinger modellen til å fokusere på å lære å klassifisere disse datapunktene i den neste iterasjonen (Bauer og Kohavi, 1999). Fokuset på å trene svakere modeller og kombinere disse, har til hensikt å føre til høyere nøyaktighet og bedre prediksjon i modellen.

Det finnes ulike metoder innenfor boosting. Den mest brukte er AdaBoost som trener en rekke svake klassifikatorer på ulike delmengder av treningsdataene. Disse vektet og kombineres til en endelig klassifikasjon (Schapire, 2013). Videre er også metoden Gradient Boosting populær. Metoden tar også utgangspunkt i en rekke svake klassifikatorer, men bruker i stedet en form for gradientoptimalisering for å justere vektet til hver modell med mål om å minimere prediksjonsfeil, i stedet for å tillegge vekt til hvert beslutningstre Friedman (2001). Videre har det kommet til ulike utvidelser av Gradient Boosting, eksempelvis XGBoost og LightGBM.

I senere tid har det blitt introdusert nye trebaserte metoder for å forbedre modellens prediksjonsevne ytterligere. Metoden XGBoost ble introdusert av Chen og Guestrin (2016), hvor Gradient Boosting tas i bruk for å forbedre prediksjonsnøyaktigheten ved å trene flere beslutningstrær sekvensielt. Metoden viser gode prediksjonsresultater i flere studier knyttet til konkursprediksjon, og et stort antall studier oppnår høyere ytelse ved anvendelse av XGBoost (Papik og Lenka, 2023). Blant disse er Zikeba mfl. (2016) med en AUC på 96%, Carmona mfl. (2019) med en AUC på 98% og Qian mfl. (2022) med en AUC på 93%.

2.4 The Black Box Problem og Explainable AI

Problemet med svarte bokser (The Black Box Problem) refererer til utfordringen med å forstå hvordan komplekse systemer kommer til sine beslutninger og resultater (Burrell, 2016). Noen typer maskinlæringsmodeller er et slikt komplekst system, og anses som vanskelig å forstå og forklare på grunn av deres kompleksitet og manglende transparens i sine interne beslutningsprosesser (Zednik, 2021). De svarte boksene gjør det vanskelig å forklare og tolke maskinlæringsmodellens handlingsmønster. Utfordringen med dette er at det er vanskelig å tolke maskinlæringsmodellen, og hvorvidt resultatene som produseres av den er pålitelige og riktige (von Eschenbach, 2021). Mangelen på innsikt i hvordan modellene tar beslutninger gjør det også vanskelig å utføre forbedringer og justeringer i modellene for å oppnå bedre resultater.

Utfordringen utspiller seg spesielt for bruksområder hvor transparens og validering er avgjørende, som for eksempel innenfor banksektoren (Zerilli mfl., 2019). Innen konkursprediksjon er dette svært sentralt i form av at det er essensielt å forstå hvorfor en bedrift blir predikert til konkurs eller ikke.

xAI er en tilnærming som forsøker å nøytralisere utfordringene knyttet til svarte bokser ved å gjøre maskinlæringsmodeller mer transparente og forståelig for mennesker (von Eschenbach, 2021). Dette omfatter blant annet modelltolkning, datavisualisering og bruk av standardiserte metoder for forklaring av modellen og dens resultater (Adadi og Berrada, 2018). Videre gir det brukerne mulighet til å kontrollere og validere modellene, som igjen bidrar til økt tillit og anvendelighet.

Formålet med bruk av xAI er å ta i bruk metoder for å løse problemet med svarte bokser, samtidig som at det ikke går på bekostning av prediksjonsevnen til maskinlæringsmodellen (Arrieta mfl., 2020). Hovedmålet er med andre ord å ta i bruk metoder for å enklere kunne forklare maskinlæringsmodellens beslutninger med hensikt å skape forståelse og tillit hos mennesker, samtidig som et høyt nivå av prediksjonsevne opprettholdes (Arrieta mfl., 2020). Dette kan gjøres ved å ta i bruk teknikker som illustrerer hvilke variabler og funksjoner som anses som viktige for å ta enkeltbeslutningen.

2.5 Tolkning

Generelt skiller litteraturen mellom to typer tolkbarhet innenfor maskinlæring, som avhenger av når tolkbarheten oppnås (Du mfl., 2019): (1) intrinsisk tolkbarhet, og (2) post hoc tolkbarhet. Intrinsisk tolkbarhet sikter til maskinlæringsmodeller som er tolkbare i seg selv, hvor modellen kan gi forklaringer på hvordan en beslutning er tatt, uten å ta i bruk spesifikke tolkningsteknikker. Denne formen for tolkbarhet kommer gjerne som et resultat av en enkel modellstruktur. Eksempler på modeller med høy intrinsisk tolkbarhet inkluderer lineære modeller som logistisk regresjon og enkle beslutningstrær, hvor beslutningsprosessen kan forklares på en enkel og intuitiv måte (Molnar mfl., 2022).

Post hoc tolkbarhet sikter derimot til evnen til å forklare hvordan en modell har tatt en beslutning, eller utført en prediksjon, i etterkant av at beslutningen eller prediksjonen er gjennomført (Lipton, 2018). Dette oppnås ved å ta i bruk tolkbarhetsmetoder etter at modellen er trent, ved at man tar i bruk en modell nummer to for å forklare den opprinnelige modellen (Du mfl., 2019). Formålet med dette er å

få innsikt i hva som påvirket modellens beslutning, og å gjøre det enklere å forstå og stole på modellens beslutninger. Innenfor post hoc tolkbarhet skiller vi mellom to typer forklaringer: global og lokal.

2.5.1 Global forklaring

En global forklaring forsøker å gi en generell oversikt over modellens handlingsmønster ved å identifisere de interne funksjonene (f.eks. variabler) som modellen anser som viktige for å foreta beslutninger (Zhou mfl., 2021). Global forklaring kan derfor gi informasjon om hvilke funksjoner som generelt er viktig for modellen, og hvordan disse påvirker beslutninger innenfor klassifiseringen. Formålet ved bruk av global forklaring er at maskinlæringsmodellen skal trenes til å lære mønstre og tendenser som fortløpende kompiles inn i modellens struktur og hyperparametere (Du mfl., 2019). Globale forklaringsmodeller gir dermed innsikt i modellens generelle handlingsmønster og logikk, og identifiserer betydningen av de interne funksjonene og hvordan disse påvirker modellens prediksjoner. Siden banker og andre finansinstitusjoner oppfattes å være tilbakeholdene med å ta i bruk komplekse konkursprediksjonsmodeller, er derfor bruk av globale forklaringsmodeller et verktøy for å gjøre modellene transparente og anvendelige i større grad.

2.5.2 Lokal forklaring

En lokal forklaring forsøker å forklare enkeltbeslutninger eller enkeltklassifiseringer i modellen (Zhou mfl., 2021). Det vil si at for en enkelt input-funksjon gir den en forklaring på hvordan modellen har foretatt en beslutning. Lokale forklaringer anses som spesielt nyttige i tilfeller der modellen har tatt en uventet beslutning, eller når brukeren ønsker forklaring på en beslutning som har blitt foretatt (Lundberg mfl., 2020). Innenfor konkursprediksjon brukes lokale forklaringsmodeller for å forstå hvordan en modell har tatt en konkret beslutning om den gjeldende observasjonen er konkurs eller ikke. Det kan derfor sies at lokal forklaring kan gi innsikt i hvilke variabler som påvirker konkurssannsynligheten for en spesifikk bedrift. Videre vil også lokale forklaringsmodeller kunne identifisere svakheter ved modellen som kan forbedre ytelsen til modellen, og dermed bidra til å redusere feilmargin og forbedre nøyaktigheten. Dette øker igjen forståelse og tillit til konkursprediksjonsmodellen, og bidrar dermed til bedre beslutningsprosesser i næringslivet.

2.5.3 Tolkbarmetoder

I litteraturen er spesielt to tolkningsmetoder mye anvendt: Local Interpretable Model-agnostic Explanations (LIME) og SHaply Additive ExPlanations (SHAP). LIME anses som en lokal post-hoc tolkbarmetode, som tar sikte på å forklare hvordan en modell foretar beslutninger. Dette gjøres gjennom lokale tolkninger av bestemte enkeltprediksjoner, ved å gjøre endringer på enkelte variabler for å se effekten på prediksjonen og dermed avklare hvor viktig variabelen er for den helhetlige prediksjonen (Molnar mfl., 2022). Det bør derimot nevnes at LIME er mindre egnet for globale forklaringer av prediksjonsmodeller (Molnar mfl., 2022). LIME-rammeverket har vist seg å være effektivt for lokal tolkning av flere maskinlæringsmetoder, hvor blant annet Park mfl. (2021) tar i bruk LIME for å forklare variabelviktighet i alle datapunktene i sine modeller med XGBoost og LightGDM. De konkluderer videre med at bruk av LIME-rammeverket evner å trekke ut variabelviktighet på en slik måte at det er mulig å observere de viktigste variablene i modellen. LIME har også vist seg effektiv i forklaring av dype nevralt nettverk, i for eksempel Chou (2019) sin studie, hvor det utvikles en hybridmodell av enkle beslutningstrær og dype nevralt nettverk for å balansere nøyaktighet og tolkbarhet mellom modellene. Resultatene av studien var prediksjonsnøyaktighet på respektive 80% og 87% for henholdsvis dype nevralt nettverk og beslutningstre, men at total nøyaktighet ble forbedret til 91% av hybridmodellen (Chou, 2019).

Schalck og Yankol-Schalck (2021) gjennomfører en studie på konkursprediksjon på franske SMB, hvor de finner at XGBoost gir god prediksjonsevne. De bruker videre SHAP for å tolke resultater og variabelviktighet. Dessuten er det gjort flere studier knyttet til bruk av SHAP i forbindelse med bedrifters kredittverdighet. Tran mfl. (2022) finner at bruk av SHAP har vist seg å være effektiv for å trekke ut hvilke variabler som påvirker prediksjonsutfallet, og er et praktisk bidrag til kredittvurderingsinstusjoner for å forutsi sannsynligheten for finansiell ustabilitet for vietnamesiske bedrifter. Bussmann mfl. (2021) tok i bruk en XGBoost-modell for å kartlegge kredittrisikoen til søreuropeiske SMB, hvor konklusjonen ved bruk av SHAP var at prediksjonsevnen kunne forbedres ved økt forståelse av hvilke variabler som påvirker kredittrisikoen. Babaei mfl. (2021) studerer også SMB, og eliminerer variabler med lave forklarende SHAP-verdier. Studien konkluderer med at prediksjonsevnen bedres med et mindre antall variabler fra regnskapet (Babaei mfl., 2021). Derimot finner Slack mfl. (2020) at man enkelt kan lage misvisende SHAP-tolkninger med forsett, og at dette er en klar utfordring knyttet til å ta i bruk SHAP som forklaringsteknikk for komplekse maskinlæringsmetoder i praksis.

2.6 Variabler

Det er vesentlig hvilke variabler som benyttes i modeller for konkursprediksjon. Det er blant annet disse som danner grunnlaget for hvor godt prediksjonsmodellen predikerer. Å identifisere variabelviktigheten er derfor essensielt. I kapittel 2.6.1 går vi derfor gjennom tidligere bruk av regnskapsbaserte variabler. Videre er det ikke alltid at alle variablene i datasettet er like viktige og skaper støy i modellen, derfor presenteres metoder for variabelseleksjon i 2.6.2.

2.6.1 Bruk av regnskapsbaserte variabler

En bedrifts finansielle situasjon påvirkes av flere forskjellige faktorer, slik som makroøkonomi, politiske beslutninger, de ansattes velvære, endrede innkjøpspriser eller etterspørsel. Som en naturlig følge av dette blir også en bedrifts konkurssannsynlighet påvirket av ulike faktorer. Hvilke variabler som er av størst viktighet kan variere betydelig mellom datautvalgene (Balcaen og Ooghe, 2006), dessuten kan dette også på samme måte variere mellom ulike land (Filipe mfl., 2016). På grunn av disse forskjellene er det vanskelig å finne et standardsett med variabler som kan brukes universelt på alle populasjoner. For å løse denne utfordringen er det vanlig å bruke regnskapsbaserte nøkkeltall som et mål på en bedrifts økonomiske situasjon (Balcaen og Ooghe, 2006). Dette er fordi at alle bedrifter gjennomfører regnskapsføring, og at dette til en viss grad er relativt standardisert. Likevel forekommer det forskjeller på selve regnskapsføringen og størrelsesdefinisjoner på bedrifter i ulike land. International Financial Reporting Standard (IFRS) hjelper til på dette området slik at regnskaper kan bli mer sammenlignbare. Dessuten, slik som i finansiell analyse, er det hensiktsmessig å transformere regnskapsvariablene om til finansielle nøkkeltall slik at dataene blir normalisert og at også bedrifter blir i større grad komparative. Dette er et vanlig prinsipp som gjøres ved konkursprediksjon (Paraschiv mfl., 2021). I tillegg fremhever Agarwal og Taffler (2008) tre styrker ved bruk av finansielle nøkkeltall: (1) ett enkelt regnskap fanger opp bedrifters samlede ytelse over tid, (2) det dobbelte bokføringsprinsippet som sikrer konsistens, og (3) lånebetingelser baseres på regnskapstall. Konkurs er ofte et resultat av dårlig økonomi over lenger tid og et regnskap vil fange opp nettopp dette.

Som tidligere nevnt i kapittel 2.2 er det flere studier som har gjennomført konkursprediksjon med utgangspunkt i regnskapsbaserte variabler, blant annet Altman (1968), Ohlson (1980), Zmijewski (1984) og Taffler (1984). I Altman (1968) kommer

det frem at det i tidligere studier identifiseres et stort antall signifikante variabler i forbindelse med konkursprediksjon. På bakgrunn av dette setter han sammen en liste med 22 potensielle verdifulle variabler. Disse variablene er tildelt fem forskjellige kategorier: likviditet, lønnsomhet, gjeldsgrad (leverage), soliditet og aktivitetsforhold. Altman (1968) finner at de variablene som best predikerer konkurs er: (1) arbeidskapital/totale eiendeler, (2) opptjent inntekt/totale eiendeler, (3) resultat før renter og skatt/totale eiendeler, (4) markedsverdi egenkapital/bokført verdi av total gjeld og (5) salg/totale eiendeler.

For Ohlson (1980) er ikke målet å forsøke å utvikle nye variabler. For enkelhetens skyld danner derfor de tre første modellene som ble estimert i studien grunnlaget for de brukte variablene videre i studien. Det ble her valgt relativt like variabler som i Altman (1968), med unntak av en størrelse-variabel. Denne målte bedriftsstørrelsen vektet etter bruttonasjonalinntekt. De andre variablene som velges er: (2) sum gjeld/totale eiendeler, (3) arbeidskapital/totale eiendeler, (4) kortsiktig gjeld/omløpsmidler, en dummy-variabel for (5) totale forpliktelse overstiger totale eiendeler, (6) nettoinntekt/forvaltningskapital, (7) driftsmidler/totale forpliktelser, en dummy-variabel for (8) positiv eller negativ nettoinntekt de to siste årene, og (9) endring i nettoinntekt. I motsetning til Altman (1968) finner Ohlson (1980) at størrelse-variabelen er statistisk signifikant med en relativt stor t -verdi.

Modellen som ble brukt i Zmijewski (1984) inneholder også regnskapsbaserte variabler, dog med færre antall enn de tidligere nevnte studiene. Denne studien inneholder kun tre regnskapsbaserte variabler som måler lønnsomhet, gjeldsgrad (leverage) og likviditet med variablene (1) nettoinntekt til forvaltningskapital, (2) total gjeld til totale eiendeler og (3) omløpsmidler til kortsiktig gjeld.

En annen kjent studie som bruker regnskapsbaserte variabler er Taffler (1984). Her brukes fire variabler som måler lønnsomhet, arbeidskapitalposisjon, finansiell risiko og likviditet. Denne modellprofilen blir brukt på britiske industribedrifter, mens de tre tidligere nevnte studiene tar for seg amerikanske industribedrifter (Altman, 1968; Ohlson, 1980; Zmijewski, 1984). Alle studiene oppnår godt resultat, noe som beviser at regnskapsbaserte variabler fint kan brukes til konkursprediksjon. Dessuten støtter Agarwal og Taffler (2007) og Agarwal og Taffler (2008) dette ved å ta i bruk modellen fra Taffler (1984) i nyere tid med regnskapsbaserte variabler. I tillegg finnes det en rekke andre konkursprediksjonstudier i nyere tid hvor slike variabler er brukt. Blant annet Paraschiv mfl. (2021) bruker 155 regnskapsbaserte variabler, hvorav 151 av de er finansielle forholdstall, to er dummy-variabler og de to siste er

bedriftenes levealder og totale eiendeler som et mål på bedriftens størrelse. Paraschiv mfl. (2021) baserer disse variablene på tidligere studier knyttet til konkursprediksjon hvor de tidligere nevnte studiene i dette kapittelet er sentrale. I tillegg nevner de å ha tatt utgangspunkt i variabelsett fra Shumway (2001), Altman og Sabato (2007), Campbell mfl. (2008), Härdle mfl. (2009), Tian mfl. (2015), Tian og Yu (2017), Kumar og Ravi (2007), Liang mfl. (2016) og SEBRA-modellen av Bernhardsen og Larsen (2007).

2.6.2 Metoder for variabelseleksjon

At resultatet av en prediksjon er tolkbart, er svært viktig for brukerne av modellen. Variabelseleksjon (feature selection) er med på å gjøre konkursprediksjonsmodellene mer tolkbare og nøyaktige (James mfl., 2013). I tillegg er det en effektiv dataforbehandlingsstrategi som gjør modellen enklere og forbedrer datautvinningsytelsen (Li mfl., 2017). Variabelseleksjon kan grovt sett deles inn i de tre kategoriene: wrapper, filter og embeddede metoder (Chandrashekar og Sahin, 2014). Felles for de alle er at de ekskluderer irrelevante variabler, unngår overtilpasning og reduserer problemer knyttet til høy dimensjonalitet i datasettet (Paraschiv mfl., 2021). Wrapper-metoder vurderer alle mulige kombinasjoner av variablene for å finne de variablene som er med på å gjøre prediksjonen best. Videre er metodene kjent for å gi god kvalitet av de valgte variablene, men til gjengjeld kreves det stor beregningskraft (Solorio-Fernandez mfl., 2016). Filtermetoder reduserer dimensjonaliteten i datasettet ved å fjerne irrelevante og/eller observasjoner og funksjoner som skaper støy. Dette gjør at prediksjonsmodellen lærer bedre og gi mer nøyaktige prediksjoner. Filtermetodene krever liten beregningskraft og har lite problemer med overtilpasning (Sanchez-Marono mfl., 2007). Metodene har likevel noen ulemper, deriblant at den ikke tar hensyn til interaksjoner mellom variablene. Eksempelvis kan enkelte variabler ha stor effekt når de kombineres med andre, mens de isolert sett fremstår som mindre viktige (James mfl., 2013).

Embeddede metoder er en trade-off mellom filter- og wrapper-metoder. Metodene bygger variabelseleksjon inn i selve modelleringen samtidig som de tar med interaksjonene med læringsalgoritmen. Dermed innehar embeddede metoder en gunstig egenskap i form av at de gjør variabelseleksjonen betraktelig mer effektivt enn wrapper-metoder, da den ikke trenger å evaluere variabelsettet iterativt (Li mfl., 2017). En mye brukt embedded metode i konkursprediksjon er Least Absolute Shrinkage and Selection Operator (LASSO) (Yildirim mfl., 2021). LASSO funge-

rer ved å bruke en straffeparameter (λ) for å unngå overtilpasning av modellen og for å forbedre prediksjonen.

λ er ikke en gitt verdi og det finnes ikke en optimal verdi for denne, og må derfor tilpasses den gjeldene modellen ved å teste den for ulike λ -verdier. λ er en absoluttverdi-basert straffeparameter som begrenser størrelsen på variablenes koeffisienter og straffer uviktige koeffisienter til å bli null. På den måten gjennomfører LASSO en seleksjon av variablene og hjelper oss med å velge ut et sett av variabler som anses som de viktigste for prediksjonen (Tibshirani, 1996).

Variabelseleksjon er en mye brukt metode til forbehandling av data i forskjellige typer maskinlærings- og datautvinningsproblemer (Li mfl., 2017). Flere tidligere sammenligningsstudier av variabelseleksjonsmetoder kommer frem til at LASSO er en av de beste og mest populære teknikkene innen banksektoren (Hautsch mfl., 2015; Barry mfl., 2021; Laborda og Ryoo, 2021). Innenfor konkursprediksjon finner Tian mfl. (2015) at LASSO har en overlegen out-of-sample prediktiv kraft. Videre bruker Pereira mfl. (2016) LASSO fordi metoden håndterer multikolinearitet godt og viser de ideelle egenskapene for å minimere den numeriske ustabiliteten som kan oppstå på grunn av overtilpasning.

De siste årene har en annen variabelseleksjonsmetode vokst frem som et populært alternativ. Xiaomao mfl. (2019) brukte SHAP til variabelseleksjon og finner at denne metoden gjør det like godt som andre populære metoder, og med enda bedre tolkbarhet. Samtidig konkluderer Marcílio og Eler (2020) med at SHAP gjør det bedre enn andre populære metoder (Marcílio og Eler, 2020). Videre finner Fryer mfl. (2021) imidlertid at SHAP er mer ustabil og i flere tilfeller gjør det dårligere enn andre metoder for variabelseleksjon. Derfor har vi i likhet med Paraschiv mfl. (2021) og andre tidligere studier på konkursprediksjon valgt å bruke LASSO til variabelseleksjon for våre logistiske regresjonsmodeller.

3 Datasett

I denne studien tar vi utgangspunkt i et datasett som inneholder alle ukonsoliderte årsregnskaper for norske bedrifter for regnskapsårene 2006-2020, som beskrevet i Wahlstrøm (2022). Dataene er hentet ut fra blant annet Brønnøysundregisteret, og består av 4 238 493 årsregnskaper (observasjoner). Hver observasjon er en bedrifts årsregnskap for et gitt regnskapsår, og omtales heretter som firmaår”. Hver observasjon i datasettet representerer med andre ord det ene unike årsregnskapet for en gitt bedrift i det gjeldende regnskapsåret. Hvert firmaår i datasettet har tilhørende regnskapsinformasjon som beskrevet i Appendiks A. Alle regnskapsposter i datasettet er komplette, så vi trenger ikke forholde oss til manglende verdier.

3.1 Konkurs som responsvariabel

I Norge reguleres konkurs for bedrifter av Konkursloven av 1984. I følge §60 igangsettes konkursbehandling når skyldneren er insolvent (Konkursloven (1984), §60). At en skyldner er insolvent betyr at vedkommende ikke evner å betale sine forpliktelser, og anses dermed som betalingsudyktig (Konkursloven (1984), §61). Videre understrekes det at insolvens ikke foreligger dersom den manglede betalingsevnen er midlertidig, eller om skyldnerens eiendeler og inntekter totalt dekker samlede forpliktelser (Konkursrådet, 2012).

I konkursprediksjon anses responsvariabelen som binær, hvor den klassifiseres som enten konkurs ($y=1$) eller ikke konkurs ($y=0$). Fra datasettet bruker vi variabelen «bankrupt_fs» som responsvariabel. Responsvariabelen er en dummyvariabel, hvor $y=1$ dersom årsregnskapet er bedriftens siste årsregnskap og bedriften er begjært konkurs. Dersom dette kriteriet ikke er oppfylt, er $y=0$, og bedriften kategoriseres som ikke konkurs (Wahlstrøm, 2022).

3.2 Variabler

I likhet med Paraschiv mfl. (2021) velger vi å bruke 155 forskjellige regnskapsbaserte variabler i denne studien. Det ubehandlede datasettet består ikke av disse variablene og vi er derfor nødt til å lage disse variablene. Dette gjøres ved å bruke de allerede eksisterende variablene i datasettet til å generere nye. 151 av variablene er forholdstall som lages ved å dividere og/eller multiplisere flere av regnskapsposte-

ne. To av variablene er dummy-variabler. Variabelen «dummy; 1 hvis total gjeld er større enn totale eiendeler» settes til verdien 1 dersom sum gjeld er større enn sum eiendeler, ellers settes verdien til 0. Variabelen «dummy; 1 dersom innskutt kapital er mindre enn total egenkapital» settes til 1 dersom sum egenkapital er større enn sum innskutt egenkapital, ellers settes verdien til 0. Variabelen «log (alder i år)» og variabelen «log (totale eiendeler)» er bedriftens levealder og sum eiendeler som begge er logaritmisk transformert.

3.3 Filtrering av data

Med utgangspunkt i studien til Paraschiv mfl. (2021), velger vi i denne studien å rette fokuset mot små- og mellomstorebedrifter (SMB). Små- og mellomstore bedrifter defineres i EU anbefaling 2003/361 som bedrifter med mindre enn 250 ansatte og omsetning på mindre enn 50 millioner euro per år, eller en samlet årsbalanse mindre enn 43 millioner euro. Med utgangspunkt i dette, velger vi å ekskludere sum eiendeler høyere enn 43 millioner euro, og omsetning på høyere enn 50 millioner euro. Videre ekskluderes alle selskapsformer som ikke er aksjeselskap (AS). Avslutningsvis har vi med utgangspunkt i Bernhardsen og Larsen (2007) og Paraschiv mfl. (2021) valgt å ekskludere svært små og inaktive bedrifter som kan bidra til ekstremverdier ved å ekskludere observasjoner hvor totale eiendeler er mindre enn 500 000 NOK.

Videre følger vi tidligere litteratur som Mansi mfl. (2012) og Paraschiv mfl. (2021), hvor vi ekskluderer observasjoner innenfor noen spesifikke industrier. Vi ekskluderer observasjoner innenfor industriene «finansielle- og forsikringsaktiviteter», «eiendomsaktiviteter», «offentlig administrasjon og forsvar», «elektrisitet- og gassforsyning» og «vannforsyning, kloakk og avfall», samt bedrifter innenfor industri tilknyttet investerings- og holdingformål. I tillegg ekskluderer vi observasjoner der industri ikke er oppgitt.

Vi inkluderer ikke firmaår fra 2020 i våre analyser. Dette skyldes koronapandemien som brøt ut i verden og dermed preget økonomien til bedriftene og i mange tilfeller kunne man se en reduksjon i inntekter og en økning i kostnader. Dette som følge av strenge restriksjoner gitt fra regjeringen. Etterhvert innførte regjeringen støtteordninger som bedriftene kunne benytte seg av (Rybalka, 2021; Regjeringen, 2022). Summen av dette gjorde at regnskapsåret 2020 naturligvis vil se annerledes ut enn de tidligere årene for mange bedrifter. Dessuten kan det tenkes at støtteordningene utsatte en potensiell konkurs for enkelte bedrifter. Papik og Lenka

(2023) skriver videre i sin studie at kriseår, slik som 2020, påvirker maskinlæringsmodeller i negativ retning med lavere nøyaktighet. I og med at vi i denne studien ikke ser på effekter av koronapandemien anser vi det som potensielt at regnskapsåret 2020 tilfører støy til modellen, og velger derfor å utelate dette regnskapsåret fra studien.

3.4 Winzorizing

Vi tar utgangspunkt i tidligere litteratur som Shumway (2001), Chava og Jarrow (2004), Tian mfl. (2015), Tian og Yu (2017), Paraschiv mfl. (2021) og gjennomfører winsorizing for å redusere effekten av ekstremverdier i datasettet. Winsorizingen gjennomføres på 1. og 99. persentil. Dette betyr at vi erstatter ekstremverdiene med nærmeste verdi innenfor persentilene. Dette har til hensikt å forbedre prediksjonsnøyaktighet i modellen, uten å eliminere data.

3.5 Behandlet datasett

I tabell 1 presenteres deskriptiv statistikk for det behandlede datasettet som brukes i studien.

Tabell 1: Deskriptiv statistikk for behandlet datasett

	Antall firmaår	Antall selskaper	Antall konkurs	Konkursfrekvens
2006	97 536	97 536	1 402	1,44%
2007	103 986	103 986	2 205	2,12%
2008	106 812	106 812	1 941	1,82%
2009	107 372	107 372	1 783	1,66%
2010	108 846	108 846	1 556	1,43%
2011	111 607	111 607	1 619	1,45%
2012	116 211	116 211	1 688	1,45%
2013	120 621	120 621	1 653	1,37%
2014	124 944	124 944	1 600	1,28%
2015	129 323	129 323	1 507	1,17%
2016	133 858	133 858	1 725	1,29%
2017	139 068	139 068	1 754	1,26%
2018	143 802	143 802	2 018	1,40%
2019	147 920	147 920	1 209	0,82%
Totalt	1 691 906		23 660	

Det behandlede datasettet som blir brukt i denne studien består av totalt 1 691 906 observasjoner og 157 variabler, hvorav 155 er forklaringsvariabler. Det er totalt 23 660 observasjoner som er klassifisert som konkurs, altså en andel på 1,4%. I Appendiks B finnes det en oversikt over variablene som inngår i det endelige datasettet.

4 Metode

Det skilles generelt mellom tre typer maskinlæringsmetoder: veiledet læring, ikke-veiledet læring og forsterket læring (James mfl., 2013). I denne masteroppgaven er fokuset rettet mot veiledet læring. Veiledet læring er at modellen trenes til å transformere inndata til utdata, gjennom å lære fra eksempler på inndata og tilhørende kjente utdata (Russell, 2010). Dette gjøres ved at modellen presenteres for et treningssett med inndata og tilhørende ønsket resultat for hver av inndataene. Deretter forsøker modellen å forutsi resultatet på egenhånd, og justerer modellens parametre for å redusere feilen mellom predikert resultat og ønsket resultat. Dette gjentas mange ganger, og modellen justerer seg gradvis som en kontinuerlig prosess for å lage nøyaktige prediksjoner for nye og ukjente inndata (Chollet, 2021). Vi skiller videre mellom to typer veiledet læring: regresjon og klassifikasjon (Sen mfl., 2020). Forskjellen mellom disse ligger i typen utdata som modellen forsøker å forutsi (Caruana og Niculescu-Mizil, 2006). Målet ved bruk av regresjon er å forutsi numeriske verdier, som for eksempel boligpriser og temperaturer. En regresjonsmodell lærer med andre ord å finne en funksjon som gir kontinuerlige utdata, basert på tilhørende inndata (Russell, 2010). Klassifisering handler derimot om å forutsi diskrete utdata verdier, som eksempel klasser eller kategorier (Kotsiantis mfl., 2007). Hvorvidt en bedrift går konkurs eller ikke er et klassifiseringsproblem, hvor utfallet er binært (konkurs eller ikke konkurs). Dette klassifiseringsproblemet står derfor sentralt i denne oppgaven.

4.1 Treningssett og testsett

Maskinlæring er en form for kunstig intelligens som gjør det mulig for datamaskinen å lære å utføre oppgaver uten å bli eksplisitt programmert (Samuel, 1959). Dette gjøres gjennom å trene en modell ved bruk av et datasett, som vanligvis deles opp i et treningssett og et testsett (James mfl., 2013). Treningssettet brukes for å trene modellen, og deretter brukes testsettet for å evaluere om modellen har lært det den er trent til. Dataene i testsettet er ukjente for modellen, og man kan derfor avdekke hvorvidt modellen har lært som ønsket (James mfl., 2013). Hvis det viser seg at modellen presterer godt på treningssettet, men dårligere på testsettet, så er dette et signal om at modellen er overtilpasset til treningssettet, og dermed ikke generaliserer godt til nye uavhengige data (James mfl., 2013). Overtilpasning er en utfordring som vanligvis oppstår når modellen blir mer kompleks, eller har mange parametre (Singh mfl., 2016). Overtilpasning er en utfordring man ønsker å unngå,

fordi det indikerer at modellen lærer mønstre og egenskaper av treningsdataene som ikke generaliserer godt til nye og uavhengige data, og dermed gir unøyaktige prediksjoner. For å unngå overtilpasning bruker vi derfor k-fold kryssvalidering og forward validation i modelltreningen.

Forward validation er en teknikk hvor treningssettet kun inneholder data som kommer før dataene i testsettet (James mfl., 2013). Ved å gjøre dette vil modellene bygges på grunnlag av forutgående data, for så å predikere fremover i tid. For å håndtere tidsseriedata, velger vi å ta utgangspunkt i praksis i Paraschiv mfl. (2021), og bruker teknikken Rolling Window for å sammenligne de ulike tidsperiodene knyttet til modellen.

I denne studien predikerer vi konkurs et år frem i tid, som følge av at det er dette som er mest vanlig i praksis, i tillegg til at Basel III også tar utgangspunkt i prediksjon ett år frem i tid. Deretter tar vi utgangspunkt i Paraschiv mfl. (2021) hvor de velger variabler, trener modell og evaluerer tilpasningen til treningsdata over fire regnskapsår. Videre evalueres ytelse på nye data i testsettet det påfølgende året. Vi bruker her en 10-fold kryssvalidering, hvor datasettet deles inn i 10 ulike perioder, hvor hver periode består av 5 år som vi kaller for fold. Av disse 5 årene, består treningssettet av de fire første årene, mens testsettet består av det siste året i perioden. For hver fold flytter vi oss et år frem, og for hver fold trenes en modell, som evalueres på både testsett og treningssett.

I appendiks C oppsummeres hele prosessen knyttet til inndelingen av trenings- og testsett. Tabellen viser også deskriptiv statistikk for det brukte datasettet, deriblant antallet observasjoner konkurs og ikke-konkurs per regnskapsår, samt konkursfrekvens per år.

4.2 Håndtering av ubalanserte data

Konkurs anses gjerne som noe som inntreffer sjeldent, hvor man heller ser en hyppigere forekomst av bedrifter som ikke går konkurs, enn bedrifter som faktisk går konkurs. Som utledet i kapittel 3.5, finner vi at det kun er 1,4% av totale firmaår i datasettet som ender i konkurs. Vi har med andre ord et ubalansert datasett, hvor andelen konkurs er betydelig lavere enn andelen «ikke konkurs». Dette kan føre til at prediksjonevnen svekkes, fordi modellen tenderer til å overprestere på den dominerende klassen, som i dette tilfellet er «ikke konkurs». Modeller vil derfor ofte ha dårligere ytelse på den mindre representerte klassen, som gir høyere sannsynlig-

het for feilklassifisering for den mindre representerte klassen (Kotsiantis mfl., 2006). Som et resultat av overrepresentasjon vil det derfor kunne oppstå uønskede skjevheter, fordi algoritmen vil tendere til å generere mer fra den dominerende klassen. Veganzones og Séverin (2018) finner i sin studie at en modells evne til å predikere konkurs reduseres ved et ubalanseforhold på 4:1 eller høyere. Ved tidligere studier, som f.eks. Zhou (2013) og Veganzones og Séverin (2018), for konkursprediksjon ser man derimot at i praksis så er ubalanseforholdet gjerne helt oppe i 1000:1.

For å gjøre bukt med utfordringene knyttet til ubalansert data kan resampling-metoder tas i bruk for å jevne ut forskjellene, ved å enten gjøre klassen for konkurs større (over-sampling) eller klassen for ikke-konkurs mindre (under-sampling). Det bør derimot nevnes at det er knyttet utfordringer også til bruk av disse metodene, da det ofte oppstår bias i prediksjonsmodellene, og modellen avviker i større grad fra realiteten i den virkelige verden. Zmijewski (1984) var den første til å påpeke problemer med å balansere data. Derfor velger flere studier innenfor konkursprediksjon, som for eksempel Shumway (2001), Tian mfl. (2015), Tian og Yu (2017) og Paraschiv mfl. (2021), å ikke ta i bruk slike resampling-metoder for håndtering av ubalanserte data. Vi velger å følge praksis i disse studiene, og balanserer derfor ikke datasettet.

4.3 Estimeringsmetoder

Det finnes flere forskjellige estimeringsmetoder som er brukt i forbindelse med konkursprediksjon. I denne studien skal vi først bruke en mye benyttet metode innenfor finanssektoren, logistisk regresjon. Deretter skal vi sammenligne med en annen metode som ikke er like mye brukt, men som nevnt i kapittel 2.5 har vist seg å gi gode resultater, nemlig XGBoost.

4.3.1 Logistisk regresjon

Logistisk regresjon (LR) er en statistisk metode som brukes til å modellere sannsynligheten for en binær responsvariabel, altså et klassifikasjonsproblem. Metoden er som nevnt i kapittel 2.2, mye brukt til både kredittvurdering og konkursprediksjon grunnet modellens evne til å gi gode prediksjoner og gode tolkbare resultater (Dong mfl., 2010). Derimot bør det nevnes at logistisk regresjon kan ha problemer med homoskedastisitet, som kan gjøre den noe mindre egnet til bruk innen finanssektoren (Lee mfl., 2006).

I denne studien anvendes logistisk regresjon for å modellere sannsynligheten for at et gitt firmaår kategoriseres som enten konkurs eller ikke konkurs. Responsvariabelen faller derfor innenfor én av to mulige kategorier: (1) konkurs ($y=1$) eller (2) ikke konkurs ($y=0$). Vektoren til de predikerte sannsynlighetene for konkurs $\hat{y} = \{\hat{y}_n\}_{n=1,\dots,N} \in [0, 1]^N$ er gitt ved (Paraschiv mfl., 2021):

$$\hat{y} = \iota \oslash (\iota + \exp(-\mathbf{X}\mathbf{w} - \iota w_0)) \quad (1)$$

hvor ι er $N \times 1$ vektor av enere, \oslash står for Hadamard-divisjon, $\mathbf{X} = \{x_{(n,i)}\}_{n=1,\dots,N,i=1,\dots,I}$ er en matrise av verdier for variablene i fra firmaår n , $\mathbf{w} = \{w_i\}_{i=1,\dots,I}$ er en vektor av koeffisienter, og w_0 er konstantleddet. For enkelthets skyld ser vi bort fra tidsindeksene (Paraschiv mfl., 2021).

Koeffisientene \mathbf{w} og w_0 er ikke kjente og må derfor estimeres. Datasettets treningssett blir brukt til denne estimeringen (James mfl., 2013), som gjøres ved å minimere den negative av logit-funksjonen $\ell(\mathbf{w}, w_0)$ som er gitt ved (Paraschiv mfl., 2021):

$$\ell(\mathbf{w}, w_0) = \sum_{n=1}^N [\mathbf{y} \odot (\mathbf{X}\mathbf{w} + \iota w_0) - \log(\iota + \exp(\mathbf{X}\mathbf{w} + \iota w_0))] \quad (2)$$

hvor $\mathbf{y} = \{y_n\}_{n=1,\dots,N} \in \{0, 1\}^N$ er vektoren med de faktisk klassifiserte konkurs eller ikke-konkurs, og \odot refererer til Hadamard-produktet. For å avgjøre hvorvidt de estimerte koeffisientene er signifikante, benyttes Wald-statistikk til å beregne z -verdier (Paraschiv mfl., 2021).

Hensikten ved estimeringen av koeffisientene er å finne de verdier som gjør at prediksjonen er nærmest mulig de faktiske firmaårene. Om hvorvidt gjeldende firmaår blir predikert til konkurs eller ikke-konkurs kan leses av utfallsverdien ved å sette inn de estimerte koeffisientene. Dette vil gi en verdi mellom 0 og 1, som indikerer den estimerte sannsynligheten for å tilhøre enten klasse 1 eller 0 (James mfl., 2013). En verdi nær 0 vil indikere en høy sannsynlighet for ikke-konkurs, og motsatt vil en verdi nær 1 indikere en høy sannsynlighet for konkurs.

4.3.2 Extreme Gradient Boosting

I denne studien tar vi også i bruk XGBoost for å trene konkursprediksjonsmodeller. XGBoost blir presentert av Chen og Guestrin (2016), som en effektiv, fleksibel og ressursbesparende ensemble maskinlæringsmetode som modellerer ikke-lineære

forhold. Metoden er en variant av Gradient Boosting som legger til rette for forbedringer innenfor både effektivitet og nøyaktighet, spesielt ved større mengder data (Dietterich, 2000). Metoden tar i bruk en kombinasjon av ulike ensemble metoder og beslutningstrær for å forbedre nøyaktigheten i modellen. Dette skjer gjennom gradvis trening av modellen ved å legge til nye beslutningstrær, hvor hensikten er å korrigere feil fra tidligere trær (Chen og Guestrin, 2016).

Den ensemble metoden forbedres gradvis ved itererende trening av dataene. Dette skjer ved at for hver iterasjon j , utledes en svak prediktor (weak learner) $f_j(X)$ ved å minimere residualene av XGBoost-modellen $f(X)^{[j-1]}$ fra forrige iterasjon $j - 1$ over ψN firmaår, tilfeldig valgt fra hver iterasjon slik (Wahlstrøm, 2023):

$$\arg \min_{f_j(X)} = \sum_{n=1}^{\psi N} [l(y_n, p(f(\mathbf{x}^n)^{[j-1]})) + g_n f_j(\mathbf{x}^n) + \frac{1}{2} h_n f_j^2(\mathbf{x}^n)] + \Omega(f_j(X)) \quad (3)$$

hvor $\psi \in (0,1]$ er et utvalgsforholdstall, $\Omega(f_j(X))$ er et straffelegg, p er sannsynligheten for konkurs, $\mathbf{x}^n = \{x\}_{i=1,\dots,I}$ er variabelverdiene for firmaår n , og

$$g_n = \delta_{f(\mathbf{x}^n)^{[j-1]}} l(y^n, p(f(\mathbf{x}^n)^{[j-1]})) \quad (4)$$

$$h_n = \delta_{f(\mathbf{x}^n)^{[j-1]}}^2 l(y^n, p(f(\mathbf{x}^n)^{[j-1]})) \quad (5)$$

er første- og andreordens gradient av den logistiske tapsfunksjonen gitt ved:

$$l(y^n, p(f(\mathbf{x}^n)^{[j-1]})) = -y^n \log(p(f(\mathbf{x}^n)^{[j-1]})) - (1 - y^n) \log(1 - p(f(\mathbf{x}^n)^{[j-1]})) \quad (6)$$

Ved å bruke et straffet beslutningstre som svak prediktor, brukes $\Omega(f_j(X))$ for å bidra til å unngå overtilpasning ved å begrense minimeringen til å favorisere enkle trær ved å definere det som (Wahlstrøm, 2023):

$$\Omega(f_j(X)) = \gamma T_j + \frac{1}{2} \lambda \|w_j\|^2 \quad (7)$$

hvor T_j er antallet noder i beslutningstreet, $\|w_j\|^2$ er verdien for nodevektene, og avslutningsvis er γ og λ straffeparametere. Etter at $f_j(X)$ er utledet, vil den legges til modellen ved:

$$f(\mathbf{X})^{[j]} = f(\mathbf{X})^{[j-1]} + \nu f_j(\mathbf{X}) \quad (8)$$

hvor $f(\mathbf{X})^{[j-1]}$ er den oppdaterte XGBoost-modellen, og $\nu \in (0,1]$ er den stegvise krympeparameteren som kontrollerer konvergeringshastigheten, som sikrer at modellen kun gjør små endringer per iterasjon. Som en kontinuerlig prosess reduseres residualene i XGBoost-modellen gradvis for hver iterasjon, og antallet totale iterasjoner

J bestemmes på forhånd. Når alle iterasjoner er gjennomført, er resultatet en endelig XGBoost-modell, $f(X)^{[J]}$, som brukes for prediksjoner «out-of-sample» (Wahlstrøm, 2023).

XGBoost er en betydelig mer kompleks metode i forhold til logistisk regresjon, og har flere hyperparametere som bestemmes på forhånd for å optimalisere modellens ytelse. Å tune hyperparametere er viktig for å forbedre ytelsen til modellene (James mfl., 2013). Det er disse verdiene som styrer hvordan modellen skal ta beslutninger. Hvis dette ikke gjøres kan modellen over- eller undertilpasse treningsdataene. Da risikerer man at modellen ikke generaliserer de faktiske dataene godt, og at modellen ikke fanger opp komplekse sammenhenger. Ofte innebærer derfor tuning av hyperparametere å teste forskjellige verdier for parameterene, for deretter å sammenligne modellenes ytelse opp mot hverandre for å finne de verdiene som gir den beste modellen (Saraswat, 2016).

I XGBoost tunes deriblant hyperparametere for antall beslutningstrær, trærnes maksimale dybde, læringshastighet, og reguleringsfaktorer for å unngå overtilpasning. For å kunne oppnå så høy ytelse som mulig på testdataene, så velger vi her å tune seks ulike hyperparametere. Optimaliseringen gjøres gjennom et grid-søk, ved bruk av 3-fold kryssvalidering, over verdiene som presenteres i tabell 2. Tuningen blir kun gjennomført på treningsdata, for å unngå lekkasje av data («data leakage») og sørge for at evalueringen blir «out-of-sample».

Tabell 2: Tuning av hyperparametere for XGBoost. Tabellen viser de utprøvde verdiene til de gitte hyperparameterene i forbindelse med grid-søk.

Hyperparameter	Grid-verdier
ν	{0.1, 0.3}
ψ	{0.5, 1}
λ	{0.0, 0.4}
γ	{1, 3}
J	{25, 100}
Maks dybde for beslutningstreet	{1, 3}

4.4 Tolkning av modeller

I Miller (2019) defineres tolkbarhet som i hvilken grad mennesker kan forstå årsaken til en beslutning. Innenfor maskinlæring defineres tolkbarhet som «bruk av ma-

skinlæringsmodeller for utvinning av relevant kunnskap om domenerelasjoner som finnes i data» (Murdoch mfl., 2019). Tolkbarhet er svært sentralt ved bruk av maskinlæring, siden mennesker generelt oppfattes som tilbakeholdne med å ta i bruk maskinlæringsmetoder som ikke er direkte tolkbare, håndterbare og pålitelige (Arrieta mfl., 2020). Ved å fokusere utelukkende på prediksjonsevne og nøyaktighet, vil modellene blir mindre transparente, noe som er i tråd med avveiningen mellom tolkning og ytelse (Dovsilovic mfl., 2018). Derimot vil økt forståelse av en modell kunne muliggjøre korrigeringer av mangler. Ved å ta hensyn til tolkbarhet ved utvikling av en maskinlæringsmodell, vil man kunne forbedre implementeringsevnen fordi (Arrieta mfl., 2020): (1) tolkbarhet bidrar til å sikre upartiskhet, og dermed oppdage og korrigere for skjevheter i treningssettet, (2) tolkbarhet gjør det enklere å oppnå robusthet ved å fokusere på potensielle motstridende distraksjoner som endrer prediksjonen, og (3) tolkbarhet kan fungere som forsikring om at det kun er relevante variabler som påvirker resultatet, som garanterer kausalitet i modellbetraktningen.

4.4.1 LASSO

LASSO er en embedded metode for variabelseleksjon. Metoden ble for alvor populær etter at Tibshirani (1996) inkluderte variabelseleksjon som en del av trening av modeller. LASSO trener modellen med utgangspunkt i alle variablene, men inkluderer et straffelegg som krymper irrelevante variabler til 0, slik at de elimineres fra modellen (Tibshirani, 1996). Ved å bruke LASSO-metoden, estimerer vi koeffisientene \mathbf{w} og w_0 ved å minimere følgende i treningssettet (Paraschiv mfl., 2021):

$$-\ell(\mathbf{w}, w_0) + \lambda \|\mathbf{w}\|_1 \tag{9}$$

der $\ell(\mathbf{w}, w_0)$ er logit-funksjonen i ligning (2), $\|\mathbf{w}\|_1$ er l_1 -regulariseringen av \mathbf{w} , og λ er en positiv hyperparameter. $\lambda \|\mathbf{w}\|_1$ utgjør et straffelegg, og når λ er tilstrekkelig stor vil straffelegget bli såpass dominant at de estimerte koeffisientene i \mathbf{w} blir 0. Dette betyr at jo høyere λ er, jo flere variabler elimineres fra modellen. Det bør derfor velges en verdi for λ som er hensiktsmessig for formålet, siden λ -verdien har stor påvirkning på det endelige resultatet av modellen.

I Lasso-regresjon brukes λ som en straffeparameter for å redusere overtilpasning i modellen. λ er en hyperparameter som må optimaliseres og tilpasses den gjeldende modellen. Tilpasningen av denne hyperparameteren skjer ved hjelp av kryssvalidering, hvor vi bruker en 3-fold kryssvalidering. Her er to av foldene treningssett, mens det siste er et valideringssett. Eksempelvis kan vi starte med å si at fold 1 og 2 er

treningssett, mens fold 3 er valideringssett. Modellen trenes da opp på disse 2 folderne med treningssett og blir til slutt evaluert på valideringssettet ved hjelp av AUC. I neste steg brukes det en annen kombinasjon av folder. Da er fold 2 og 3 treningssett, mens fold 1 er valideringssett. På nytt evalueres modellen ved AUC. Til slutt er fold 1 og 3 treningssett, og fold 2 er valideringssett. Vi har da kjørt gjennom de 3 valideringssettene og den gjennomsnittlige AUC for alle valideringssettene blir da beregnet. Den verdien av λ som gir den laveste gjennomsnittlige feilen, kan velges som den optimale verdien (James mfl., 2013). Vi velger å bruke den største verdien av λ slik at den kryssvaliderte feilen er innenfor én standardfeil av den minste gjennomsnittlige kryssvaliderte feilen. Vi illustrer de ulike lambda-verdiene for hver fold i Appendix D, hvor vi ser gitt AUC mot $\log(\lambda)$. Den vertikale linjen helt til venstre gir uttrykk for den λ -verdi som gir høyest gjennomsnittlig AUC-verdi på valideringssettet.

4.4.2 SHAP

SHAP ble først utledet av Lundberg og Lee (2017), som presenterer rammeverket som et enhetlig rammeverk basert på seks eksisterende metoder. Rammeverket har til hensikt å tolke prediksjoner av komplekse modeller, ved for eksempel bruk av dyplæring eller ensemble metoder. SHAP er basert på spillteori, og tildeler hver variabel en verdi for viktighet knyttet til en bestemt prediksjon. Dette gjør det mulig å tolke variabelviktigheten, altså hvilke variabler som påvirker prediksjonen i størst grad. Til forskjell fra metoden LIME, så evner SHAP-rammeverket å gi både globale forklaringer av prediksjonsmodeller og lokale forklaringer av enkeltprediksjoner. Vi bruker derfor SHAP-rammeverket for å tolke og forklare prediksjonene fra XGBoost-modellene og deres prediksjoner. SHAP forklarer prediksjonen for et firmaår n med følgende lineære funksjon (Wahlstrøm, 2023):

$$f(\mathbf{x}^n) = \phi_0^n + \sum_{i=1}^I \phi_i^n \quad (10)$$

hvor $\phi_0^n = E[f(X)] = \frac{1}{N} \sum_{n=1}^N f(x^n) \in \mathbb{R}$ er den gjennomsnittlige prediksjonen over alle N antall firmaår, og ϕ_i^n er SHAP-verdien knyttet til prediktoren i . Disse SHAP-verdiene gjør det mulig å tolke variablenes innvirkning på prediksjonen, ved å konstatere variabelinnvirkningens omfang og retning. SHAP-verdien kan derfor på mange måter sammenlignes med variabelkoeffisienten w_i i logistisk regresjon, med unntak av at SHAP-verdien er knyttet til hvert enkelt firmaår (Wahlstrøm, 2023).

Videre beregnes SHAP-verdien til en variabel i i en prediksjon for firmaår n , som den vektete summen av det marginale bidraget for prediksjonen ved å velge å inkludere eller ekskludere variabelen fra alle mulige delsett av de andre variablene slik (Wahlstrøm, 2023):

$$\phi_i^n = \sum_{S \subseteq \mathbf{x}_i^n} \frac{|S|!(\mathbf{I} - |S| - 1)!}{\mathbf{I}!} [f(S \cup x_i^n) - f(S)] \quad (11)$$

der \mathbf{x}_i^n representerer delsett av \mathbf{x}^n som ekskluderer x_i^n , S representerer et delsett av \mathbf{x}_i^n , og $|S|$ er antall variabler i S . SHAP-verdien representerer derfor en bestemt variablers innflytelse på modellen, altså variabelviktigheten. Jo høyere SHAP-verdien er, jo høyere er variabelviktigheten (Lundberg og Lee, 2017).

Videre anvender vi SHAP beeswarm-plott for hver fold for å analysere den globale variabelviktigheten. Disse ses i Appendiks E. Her presenteres de 10 viktigste variablene i hver fold, med tilhørende SHAP-verdier per firmaår. SHAP-verdiene representerer en gitt variablers gjennomsnittlige bidrag til prediksjonen konkurs, hvor høye SHAP-verdier som nevnt over representerer stor innflytelse på prediksjonen. Beeswarm-plottet presenteres som et plott bestående av prikker, hvor hver prikk representerer et firmaår fra datasettet. Prikkens plassering på x-aksen bestemmes av SHAP-verdien, og fargene på de ulike prikkene representerer variabelens verdi (Lundberg, 2018a). Plottene analyseres videre i kapittel 5.3.

Avslutningsvis analyserer vi den lokale variabelviktigheten ved bruk av SHAP waterfall plots. Disse finnes i kapittel 5.3.2. Her er de mest betydningsfulle variablene sortert synkende basert på deres absolutte SHAP-verdi, slik at variabelen som står øverst er av størst viktighet for prediksjonen. Den gjennomsnittlige predikerte log-oddsen $E[f(\mathbf{X})]$ for modellen er vist nederst i plottet (Lundberg, 2018b). Videre representerer tallet inne i de blå og røde boksene de absolutte SHAP-verdiene. Både de blå og røde boksene trekker prediksjonen bort fra den gjennomsnittlige log-oddsen $E[f(\mathbf{X})]$, men i forskjellig retning. De blå boksene reduserer konkurssannsynligheten, mens de rød øker konkurssannsynligheten. Øverst i plottet presenteres $f(\mathbf{x})$ som representerer den predikerte log-oddsen til den gitte prediksjonen (Lundberg, 2018b).

4.5 Evaluering av modell

En sentral del av prosessen med å trene en maskinlæringsmodell er å evaluere hvor godt modellen presterer. En slik evaluering danner også grunnlag for å sammenligne

modeller med hverandre. I tidligere litteratur, som eksempelvis i Paraschiv mfl. (2021) og Tian og Yu (2017) er det to evalueringsmål som hyppig brukes: AUC og Brier score.

4.5.1 AUC

AUC (Area under the receiver operating characteristic curve) brukes som evalueringskriterie for å vurdere en modells prediksjonsevne for binære klassifiseringsproblemer, og anses i mange tilfeller å foretrekke fordi man i større grad unngår skjeve estimater (Brzezinski og Stefanowski, 2017). Disse skjeve estimatene har lett for å bli gjeldene når man benytter treffsikkerhet og andre mål som er avhengig av en terskelverdi, da disse verdiene i stor grad kan påvirke evalueringsmålet. Ved binære klassifiseringsproblemer forsøker modellen å skille mellom klasser, som i denne studien er mellom konkurs og ikke-konkurs. AUC måler hvor godt modellen klarer å skille mellom disse to klassene, ved å beregne arealet under ROC-kurven (Receiver Operating Characteristic). ROC-kurven viser plottet mellom sann positivrate og falsk positivrate ved terskelverdier fra 0 til 1 for klassifisering av prediksjonene til 0 og 1 (James mfl., 2013). Sann positivraten representerer andelen positive observasjoner som modellen identifiserer korrekt, og falsk positivraten viser andelen falske positive observasjoner som blir identifisert av modellen. AUC er en verdi mellom 0.5 og 1, hvor en verdi lik 0.5 indikerer at modellen er like god som tilfeldig gjetting, mens en verdi lik 1 indikerer at modellen kan skille positive og negative tilfeller perfekt. Vi tar utgangspunkt i Hosmer Jr. mfl. (2013) som omtaler $AUC \in [0.7, 0.8)$ som en akseptabel modell, samt $AUC \in [0.8, 0.9)$ som utmerket og $AUC \geq 0.9$ som enestående. Altså, jo høyere AUC-verdi, jo bedre evner modellen å skille mellom konkurs og ikke-konkurs.

4.5.2 Brier score

Brier score anvendes også for å evaluere prediksjon av binære klassifiseringsproblemer. Ved å følge Tian mfl. (2015), Tian og Yu (2017) og Paraschiv mfl. (2021) bruker vi Brier (1950) score som evalueringskriterie for modellens prediksjonsevne. Brier score måler avstand mellom modellens prediksjon og faktiske utfall, og gir derav en kvantitativ vurdering av hvordan modellen presterer (Brier, 1950). Brier score defineres som gjennomsnittet av kvadratene til avvikene mellom faktiske utfall

og prediksjonene (Brier, 1950):

$$\text{Brier score} = \frac{\sum_{n=1}^N (\hat{y}_n - y_n)^2}{N} \quad (12)$$

hvor N er antallet firmaår i treningssettet, \hat{y}_n er predikert sannsynlighet for konkurs i firmaår n , og y_n er faktisk klassifisert konkurs eller ikke-konkurs for firmaår n . Brier score er en verdi mellom 0 og 1. Hvis prediksjonen samsvarer nøyaktig med faktisk utfall, vil Brier score være lik 0. Er prediksjonen feil er Brier score lik 1. Modellens prediksjon evalueres dermed som mer nøyaktig, jo lavere Brier score er (Paraschiv mfl., 2021).

5 Resultat og diskusjon

I dette kapitlet presenterer vi resultater ved bruk av maskinlæringsmetodene som utledes i kapittel 4. Først viser vi i delkapittel 5.1 hvilke variabler som velges av LASSO og dermed benyttes ved prediksjon med logistisk regresjon. Deretter sammenlignes prestasjonen til modellene med modeller som benytter XGBoost, og finner at XGBoost gir bedre prediksjoner i delkapittel 5.2. Avslutningsvis presenterer vi i delkapittel 5.3 hvordan XGBoost kan tolkes med SHAP.

5.1 Variabelseleksjon ved LASSO

Appendiks D presenterer LASSO stiplott for hvert av de 10 foldene med alle firmaår, for henholdsvis 2010-2019 som testår. Hvert stiplott viser verdier av λ fra en verdi som er så høy at alle estimerte koeffisienter settes til 0 (altså at ingen variabler blir valgt) til den verdien av λ som blir endelig valgt for hver fold (se beskrivelse i kapittel 4.4.1). Under hvert stiplott er et tilhørende plott som viser AUC ved forskjellige verdier av λ . Med reduserende verdi av λ viser plottene i Appendiks D at flere variabler blir valgt og at AUC øker. Aksen over plottene av AUC viser antall valgte variabler for hver verdi av λ . For å få en robust modell ønsker vi å utelukke støy i form av variabler med svak forklaringskraft, samtidig som at dette ikke reduserer modellens prediksjonsevne ved at AUC reduseres. I stiplottene anses variabler som inkluderes ved en høyere verdi av λ som å ha relativt høyere variabelviktighet. I AUC-plottene for hver fold reduseres AUC minimalt frem til det gjenstår 10 variabler. Vi velger derfor å ta i bruk de 10 viktigste variablene for hver fold, i tråd med praksis i blant annet Paraschiv mfl. (2021).

Tabell 3: Tabellen viser estimerte regresjonskoeffisienter og z-verdier (i parentes) ved logistisk regresjon til foldenes ti viktigste variabler som LASSO har valgt ut.

Variabel	Kategori	2010	2011	2012	2013	2014
Konstantledd		-4.05 (-106.65)	-4.17 (-115.35)	-4.30 (-111.87)	-4.22 (-100.26)	-4.26 (-97.79)
Dummy; 1 hvis total gjeld er større enn totale eiendeler	Leverage	0.41 (11.24)	0.23 (6.01)	0.24 (6.17)	0.41 (10.45)	0.43 (10.46)
Leverandørgjeld/totalte eiendeler	Likviditet	1.65 (28.17)	1.63 (28.27)	1.65 (27.81)	1.76 (28.25)	1.77 (28.32)
Årsresultat/totalte eiendeler	Lønnsomhet	-0.05 (-0.26)	-0.78 (-17.39)	-0.75 (-16.56)	-0.89 (-18.15)	-0.89 (-17.81)
Dummy; 1 hvis innskutt kapital er mindre enn total egenkapital	Soliditet	-0.86 (-23.33)	-0.81 (-22.40)	-0.82 (-21.45)	-0.72 (-18.53)	-0.66 (-16.50)
(kortsiktig gjeld – kortsiktig likviditet)/totalte eiendeler	Leverage	0.27 (6.98)	0.53 (12.95)	0.48 (11.71)	0.10 (2.37)	0.14 (3.10)
Rentekostnader/totalte eiendeler	Soliditet	10.28 (22.78)	10.35 (23.55)	10.54 (23.15)	11.60 (23.64)	11.73 (22.55)
Skyldig offentlige avgifter/totalte eiendeler	Likviditet	3.58 (29.93)	3.46 (29.67)	3.55 (29.91)	3.89 (31.91)	4.02 (33.71)
Log (alder i år)	Alder	-0.36 (-31.90)	-0.38 (-32.96)	-0.37 (-30.94)	-0.39 (-32.57)	-0.41 (-34.01)
Inventar/omløpsmidler	Likviditet	0.73 (18.00)	0.71 (17.58)	0.68 (16.32)	0.64 (14.92)	0.56 (12.81)
Resultat for skatt/totalte eiendeler	Lønnsomhet	-0.88 (-4.95)				
Totalte eiendeler/(total egenkapital + langsiktig gjeld)	Leverage		-0.25 (-11.97)	-0.21 (-9.22)		
Kortsiktig likviditet/omløpsmidler	Likviditet				0.64 (14.92)	-1.00 (-15.52)
AUC		0.87	0.87	0.87	0.87	0.87
Brier Score		0.0133	0.0133	0.0133	0.0127	0.0119

Variabel	Kategori	2015	2016	2017	2018	2019
Konstantledd		-4.24 (-95.18)	-4.22 (-91.02)	-4.07 (-93.74)	-4.04 (-93.84)	-4.03 (-96.27)
Dummy; 1 hvis total gjeld er større enn totale eiendeler	Leverage	0.42 (10.15)	0.45 (10.51)	0.31 (6.80)	0.32 (7.21)	-1.34 (-21.64)
Leverandørgjeld/totalte eiendeler	Likviditet	1.78 (28.60)	1.66 (26.74)	1.74 (28.49)	1.74 (28.80)	1.70 (29.26)
Årsresultat/totalte eiendeler	Lønnsomhet	-0.87 (-17.18)	-0.84 (-16.41)	-0.76 (-15.34)	-0.74 (-15.30)	-0.69 (-15.17)
Dummy; 1 hvis innskutt kapital er mindre enn total egenkapital	Soliditet	-0.60 (-15.04)	-0.50 (-12.24)	-0.45 (-11.10)	-0.44 (-10.99)	-0.47 (-12.13)
(kortsiktig gjeld – kortsiktig likviditet)/totalte eiendeler	Leverage	0.15(3.54)	0.18 (4.06)	0.39 (8.10)	0.37 (7.77)	0.38 (8.59)
Rentekostnader/totalte eiendeler	Soliditet	10.78 (20.08)	10.65 (19.05)	10.65 (18.37)	10.33 (17.15)	10.94 (18.01)
Skyldig offentlige avgifter/totalte eiendeler	Likviditet	4.41 (38.20)	4.55 (40.19)	4.54 (41.80)	4.57 (42.65)	4.46 (43.37)
Log (alder i år)	Alder	-0.42 (-35.86)	-0.43 (-36.56)	-0.44 (-37.64)	-0.44 (-37.91)	-0.43 (-37.25)
Inventar/omløpsmidler	Likviditet	0.55 (12.45)	0.47 (10.57)			
Resultat for skatt/totalte eiendeler	Lønnsomhet					
Totalte eiendeler/(total egenkapital + langsiktig gjeld)	Leverage			-0.19 (-9.89)	-0.18 (-9.20)	-0.17 (-9.35)
Kortsiktig likviditet/omløpsmidler	Likviditet	-1.12 (-17.49)	-1.29 (-19.94)	-1.27 (-20.38)	-1.34 (-21.64)	-1.23 (-21.13)
AUC		0.88	0.88	0.87	0.87	0.87
Brier score		0.0110	0.0120	0.0117	0.0130	0.0079

I tabell 3 presenteres de 10 viktigste variablene for hver fold, samt deres tilhørende regresjonskoeffisienter og z-verdier. Ved å se på variablenes z-verdier, ser vi at av de åtte variablene som går igjen i alle foldene, så er det kun variabelen «årsresultat/total eierandel» som ikke er signifikant på 5%-nivå på tvers av alle foldene. Denne variabelen er signifikant i alle folder unntatt fold 1, hvor den har en relativt beskjeden koeffisientverdi. Avslutningsvis utmerker variabelen «rentekostnader/total eierandel» seg spesielt med relativt høyere koeffisientverdier, i tillegg til å ha svært signifikante z-verdier. På tvers av alle de 10 foldene, er det variablene «dummy; 1 hvis gjeld > total kapital» og «leverandørgjeld/total kapital» som utmerker seg som de to viktigste variablene i stiplottene. I tillegg er variabelen «årsresultat/total eierandel» tredje viktigste variabel på tvers av foldene, til tross for at denne ikke er signifikant i første fold.

Ved LASSO stiplott og AUC-plott presenteres variabelseleksjon ved bruk av LASSO på en oversiktlig måte. I Appendix D presenteres hvordan de ti viktigste variablene påvirker modellens AUC. Fremstillingen er hensiktsmessig for å se sammenhengen mellom antall utvalgte variabler og hvor god prediksjonsevne modellen har. Sett opp mot tidligere studier som også benytter LASSO ved logistisk regresjon, eksempelvis Paraschiv mfl. (2021), er den generelle oppfatningen at LASSO fungerer godt for variabelseleksjon for logistiske regresjonsmodeller.

5.2 Modellestimering

Tabell 4 viser en oversikt over verdier for AUC og Brier score, for de to ulike maskinlæringsmetodene logistisk regresjon og XGBoost, på tvers av alle folder. Logistisk regresjon benytter variablene som velges av LASSO, som utledet i kapittel 5.1. For XGBoost benyttes alle variabler, da XGBoost som forklart i kapittel 4.3.2 bruker en kombinasjon av ulike ensemble metoder og beslutningstrær, og dermed evner å håndtere en større mengde variabler uten at det oppstår overtilpasning (Chen og Guestrin, 2016). Videre har både AUC og Brier score til hensikt å måle modellens prediksjonsevne. Vi presenterer verdier for både treningssett (in sample) og testsett (out of sample).

Tabell 4: Modellenes prediksjonsevne i AUC og Brier-verdier for hver fold i trenings- og testsett.

Fold	AUC Treningssett		AUC Testsett		Brier Treningssett		Brier Testsett	
	LR	XGBoost	LR	XGBoost	LR	XGBoost	LR	XGBoost
1	0.86	0.91	0.87	0.89	0.0162	0.0153	0.0133	0.0131
2	0.86	0.90	0.87	0.89	0.0162	0.0156	0.0133	0.0131
3	0.87	0.90	0.87	0.89	0.0147	0.0129	0.0133	0.0131
4	0.87	0.92	0.87	0.89	0.0138	0.0129	0.0127	0.0125
5	0.87	0.92	0.87	0.90	0.0131	0.0124	0.0119	0.0116
6	0.87	0.92	0.88	0.90	0.0128	0.0125	0.0110	0.0108
7	0.87	0.92	0.88	0.90	0.0122	0.0115	0.0120	0.0117
8	0.88	0.92	0.87	0.90	0.0119	0.0112	0.0117	0.0115
9	0.88	0.92	0.87	0.90	0.0117	0.0110	0.0130	0.0127
10	0.87	0.92	0.87	0.90	0.0119	0.0112	0.0079	0.0078

5.2.1 Logistisk regresjon

I de logistiske regresjonsmodellene ligger AUC-verdiene mellom 0.86 og 0.88 på treningssettet, og mellom 0.87 og 0.88 på testsettet. Basert på vurderingskriteriene presentert i kapittel 4.5.1, ligger AUC på tvers av foldene innenfor $[0.8, 0.9)$, og vurderes dermed som utmerket, både for treningssett og testsett.

Videre presenteres de logistiske regresjonsmodellenes Brier score, hvor verdiene for treningssett ligger mellom 0.0117 og 0.0162, og mellom 0.0079 og 0.133 på testsettet. Basert på vurderingskriteriene presentert i kapittel 4.5.2, anser vi Brier verdiene som relativt lave både på treningssett og testsett, da disse er nær 0. Vi ser også en tendens til at Brier-verdiene er lavere for test-sett, som indikerer god prediksjonsevne. Dette samsvarer i stor grad med AUC-verdiene for modellene.

5.2.2 XGBoost

For treningssettene på XGBoost-modellene ser vi at AUC-verdiene ligger mellom 0.90 og 0.92. Mens for testsettene ligger verdiene mellom 0.89 og 0.90. Alle AUC-verdiene er lavere for testsettene enn for treningssettene. Basert på de tidligere nevnte vurderingskriterier ser vi at modellene får status som utmerket på treningssettene. På testsettene vurderes fire av modellene som utmerket, samtidig som de seks resterende modellene vurderes til enestående.

For XGBoost-modellene ligger brier-verdiene mellom 0.0110 og 0.0156 på treningssettene. På testsettene varierer verdiene mellom 0.0078 og 0.0131. For seks av modellene er verdiene på testsettene lavere enn for treningssettene. Og motsatt, for fire

av modellene er verdiene for testsettene høyere enn for treningssettene.

5.2.3 Sammenligning av logistisk regresjon og XGBoost

Først og fremst finner vi at XGBoost gjør det bedre enn logistisk regresjon for samtlige perioder ved begge evalueringsmålene. Dette er i tråd med det vi presenterer i kapittel 2.1.1 om trade-offen mellom tolkning og ytelse, hvor det presenteres at mer komplekse maskinlæringsmodeller som regel har bedre prediksjonsevne.

Våre logistiske regresjonsmodeller tenderer til å gjøre det bedre på enkelte testsett i forhold til treningssett. Dette er i tråd med Tian mfl. (2015) sin studie, hvor de finner at LASSO påvirker modellen slik at den får svært god prediktiv kraft out of sample. Samtidig gjør XGBoost det jevnt over bedre på sine treningssett. Det kan tenkes at en mulig årsak til dette er at LASSO trekker ut de viktigste variablene før selve estimeringen skjer i logistisk regresjon. På den måten gjennomføres estimeringen kun med de utvalgte variablene og kan være av stor betydning for selve modellestimeringen, noe som også kommer frem i studien til Paraschiv mfl. (2021). I tillegg kan dette tyde på en viss grad av overtilpasning, som i kapittel 4.1 nevnes som et kjent fenomen knyttet til komplekse maskinlæringsmetoder.

Siden Trinkle og Baldwin (2007) finner at selv en liten økning i AUC vil ha stor påvirkning på bankers lønnsomhet, vil det faktum at XGBoost gjør det bedre enn logistisk regresjon være et sterkt argument for implementering av den mer komplekse maskinlæringsmetoden. Den økte lønnsomheten oppstår i forbindelse med at kostnadene som oppstår ved feilprediksjon av konkurs reduseres. Banker som låner ut penger og velger en XGBoost-modell for konkursprediksjon vil da få en større fordel sammenlignet med banker som velger en logistisk regresjonsmodell. Dette er fordi at bankene som anvender XGBoost-modeller mest sannsynlig vil låne ut til færre bedrifter som kommer til å gå konkurs. Dermed blir det potensielle tapet mindre og lønnsomheten bedre. En XGBoost-modell kan på den måten redusere det potensielle tapet til banken. For kostnader knyttet til at banken predikerer en ikke konkurs bedrift til å være konkurs, vil økt nøyaktighet potensielt gi økt lønnsomhet fordi banken låner ut til flere gode kunder, og derfor oppnår økte renteinntekter.

Avslutningsvis kan økt prediksjonsevne ved XGBoost skyldes optimalisering av hyperparameterne som er gjennomført. Det kan tenkes at om optimaliseringen er gjennomført på en hensiktsmessig måte så kan hyperparameterne påvirke det endelige resultatet i positiv grad. Dette står i motsetning til logistisk regresjon hvor LASSO

kun krever tilpasning av λ . Det kan derfor tenkes at XGBoost-modellene i større grad blir påvirket av valgene gjort under optimaliseringen enn det de logistiske regresjonsmodellene gjør. Dette er et fenomen som blant annet anses som en valid forklaring i studien til Xia mfl. (2017).

5.3 SHAP

Vi finner at XGBoost predikerer bedre enn logistisk regresjon, men det er knyttet en utfordringer til at XGBoost er vanskeligere å tolke. Vi skal derfor i dette kapittelet forsøke å vise hvordan dette problemet kan løses ved bruk av SHAP. Som nevnt i kapittel 4.4.2 bruker vi globale og lokale forklaringer for å forsøke å tolke XGBoost-modellene. SHAP-rammeverket presenteres derfor her som en mulig løsning for den manglende tolkbarheten som følger med komplekse modeller.

5.3.1 Global forklaring

I Appendiks E presenteres SHAP Beeswarm plott for hvert testår i perioden 2010 til 2019. Disse plottene viser hvilke variabler som har høyest absolutt SHAP-verdi, og dermed høyest variabelviktighet for XGBoost-modellene. Her presenteres også hvilke effekter disse variablene har på modellen. Videre presenteres tilhørende standardiserte SHAP-verdier i tabell 5. Disse gir et innblikk i de ulike variablenes viktighet for prediksjonene i hver av XGBoost-modellene. SHAP-verdier tolkes, som nevnt i kapittel 4.4.2, slik at en høy verdi indikerer høy variabelviktighet. Vi har plukket ut de 25 variablene med høyest SHAP-verdier på tvers av alle testårene. Den variabelen med høyest viktighet per testår er satt til verdien 100, og deretter settes SHAP-verdiene til resterende variabler i forhold til dette.

Tabell 5: Standardiserte SHAP-verdier

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Gjennomsnitt
omløpsmidler/kortsiktig gjeld	5,739	7,710	13,462	17,425	32,566	22,974	20,947	10,642	12,480	7,637	15,158
kortsiktig likviditet/salg	19,582	0,000	14,239	18,640	0,000	29,319	34,878	49,654	41,364	54,380	26,206
kortsiktig likviditet /kortsiktig gjeld	12,905	20,369	15,211	6,673	31,212	10,100	35,395	21,247	33,620	36,303	22,303
effektive skattesats	37,749	21,334	16,631	32,456	30,261	32,321	23,284	31,317	39,517	42,198	30,707
totale egenkapital/totale eiendeler	37,444	35,708	27,761	20,508	0,844	6,100	0,793	1,880	5,230	8,409	14,468
årsresultat/salg	44,124	19,645	5,318	12,563	0,152	0,371	0,153	0,527	10,074	3,332	9,626
total egenkapital/salg	11,131	30,505	28,959	24,724	34,178	26,360	23,988	25,333	35,565	31,481	27,222
opptjent egenkapital/salg	32,846	31,947	61,320	42,851	25,046	10,638	12,180	11,606	9,216	12,648	25,030
årsresultat/totale eiendeler	12,109	16,991	23,910	38,575	37,165	12,365	16,353	40,183	9,756	22,359	22,976
leverandørgjeld/totale eiendeler	62,281	88,104	87,512	67,629	56,555	44,660	37,215	37,753	66,276	45,428	59,341
skyldig offentlige avgifter/totale eiendeler	88,980	88,924	76,606	100,000	91,964	98,761	100,000	100,000	95,385	98,130	93,875
varekostnad/inventar	12,407	4,105	0,000	20,691	26,860	35,839	32,727	29,850	19,056	20,494	20,203
varekostnad/salg	14,731	21,549	21,112	30,029	29,684	14,357	8,487	10,294	16,882	14,536	18,166
utbytte/årsresultat	76,157	76,688	61,275	60,630	50,071	35,385	32,645	26,985	35,687	37,845	49,338
renteutgifter/total gjeld	17,829	25,370	25,077	33,633	31,554	23,994	14,262	15,204	19,379	12,179	21,848
renteinntekter/renteutgifter	55,329	90,089	79,795	57,365	63,852	48,155	51,087	53,329	45,207	47,040	59,125
renteinntekter/totale eiendeler	48,999	35,421	37,534	42,504	39,830	39,647	28,312	27,798	29,893	28,656	35,859
soliditetsgrad	34,815	28,565	30,765	38,193	21,925	33,372	40,618	26,276	52,742	31,194	33,846
(kortsiktig gjeld - kortsiktig likviditet)/totale eiendeler	38,154	55,604	53,660	48,409	52,661	60,525	80,250	85,447	87,442	62,043	62,419
leverandørgjeld/salg	68,372	75,130	65,856	53,727	48,687	49,619	49,774	43,147	42,330	59,078	55,572
opptjent egenkapital/kortsiktig gjeld	49,088	77,695	58,171	20,553	25,611	19,384	19,868	25,535	36,390	35,791	36,809
(total egenkapital - immaterielle eiendeler)/(totale eiendeler)	74,741	100,000	86,645	82,896	91,381	72,615	56,520	33,513	21,436	45,305	66,505
resultat for skatt/kortsiktig gjeld	30,739	13,361	10,728	9,595	18,897	23,783	25,720	23,606	6,195	18,702	18,133
log (alder i år)	100,000	97,340	100,000	94,572	100,000	100,000	92,207	91,783	100,000	100,000	97,590
log (totale eiendeler)	41,664	29,820	38,131	31,067	34,631	42,072	48,024	69,338	76,496	76,879	48,812

Først og fremst indikerer tabell 5 at variabelen « $\log(\text{alder i år})$ » er den variabelen som skiller seg spesielt ut med høyest SHAP-verdier på tvers av alle testårene. Denne variabelen anses som den viktigste i hele seks av ti testår. Videre fremstår variabelen «skyldig offentlige avgifter/total eiedeler» med jevnt over høye SHAP-verdier, hvorav den er av høyest viktighet i tre av ti testår. Variabelen « $(\text{total egenkapital} - \text{immaterielle eiendeler})/(\text{total eiedeler} - \text{immaterielle eiendeler})$ », er den variabelen med høyest viktighet i det resterende teståret.

I Beeswarm plottene i Appendiks E observerer vi at en lav variabelverdi for variabelen « $\log(\text{alder i år})$ » påvirker prediksjonen mot høyere sannsynlighet for konkurs. For variabelen «skyldig offentlige avgifter/total eiedeler» ser vi motsatt at en høy variabelverdi tenderer å gi høyere sannsynlighet for konkurs. For variabelen « $(\text{total egenkapital} - \text{immaterielle eiendeler})/(\text{total eiedeler} - \text{immaterielle eiendeler})$ » tenderer lave verdier til å gi høyere sannsynlighet for konkurs. Disse variablene har høyest standardiserte SHAP-verdier på tvers av alle foldene.

Det er videre sentralt å undersøke hvorvidt funnene som SHAP-rammeverket presenterer er pålitelige, og om den relative økningen i tolkbarhet gir grunnlag for å ta i bruk mer komplekse modeller i større grad i finanssektoren i dag. Dette gjør vi ved å sammenligne variabeleffektene fra modellene med økonomisk teori og tidligere litteratur, samt en viss grad sunn fornuft og intuisjon. Vi undersøker derfor de ti variablene med høyest gjennomsnittlig viktighet på tvers av alle foldene.

Tabell 5 indikerer at « $\log(\text{alder})$ » er den variabelen som er viktigst for å predikere konkurs i våre modeller. Av beeswarm plottene kommer det frem at en lav verdi for $\log(\text{alder})$ øker sannsynligheten for konkurs. Succurro mfl. (2014) hevder at det er en negativ korrelasjon mellom bedriftens alder og konkurssannsynlighet. Dette indikerer med andre ord at jo høyere bedriftens alder er, jo lavere er sannsynligheten for konkurs. Dette anses som naturlig, da eldre bedrifter gjerne har opparbeidet seg en solid økonomisk posisjon, samt godt omdømme i markedet. Yngre bedrifter har derimot vanligvis mindre kapital, høyere gjeldsnivåer og lavere inntekter. Dette er gjerne en rot til større utfordringer med håndtering av økonomisk usikkerhet. Videre indikerer også funn fra blant annet Altman mfl. (1977), Taffler (1983) og Vallini mfl. (2008) at yngre bedrifter tenderer til å ha høyere sannsynlighet for konkurs. Det er derfor naturlig å antyde at dette funnet fra modellen er pålitelig.

Videre vurderes «skyldig offentlige avgifter/total eiedeler» til å ha høy variabelviktighet for samtlige modeller i perioden, og at høye verdier av denne variabelen tenderer til å øke konkurssannsynligheten. Nøkkeltallet representerer hvor mye en

bedrift skylder staten i skatter og avgifter, og er derfor en indikasjon på en bedrifts likviditet. Denne variabelen kan være spesielt viktig fordi skattemyndigheten er en hard inndriver av penger, og er dermed den første kreditoren bedrifter prioriterer å betale det de skylder (Eklund mfl., 2001). Derfor er det faktum at bedrifter ikke kan betale sine skatteforpliktelser en sterk indikator på likviditetsproblemer. I tidligere litteratur er det bred enighet om at det er negativ korrelasjon mellom likviditet og økt konkurssannsynlighet. Dette er i tråd med allmenn økonomisk teori knyttet til at god likviditet er en viktig faktor for å unngå konkurs (Hicks, 1962). Funnene fra modellen er derfor i stor grad i tråd med økonomisk teori og sunn fornuft.

Variabelen $\ll(\text{total egenkapital-immaterielle eiendeler})/(\text{totale eiendeler-immaterielle eiendeler})\gg$ er også en variabel som er av spesiell viktighet for flere modeller, til tross for betydelige lavere viktighet i årene 2017 og 2018. Variabelen omtales vanligvis som den materielle egenkapitalandelen (TCE), og gir en indikasjon på bedriftens økonomiske situasjon basert på dens materielle eiendeler, og hvorvidt bedriften evner å håndtere potensielle tap. Beeswarmplottene for modellene viser at en lav verdi av variabelen øker konkurssannsynligheten. En høy verdi av denne variabelen indikerer god likviditet og evne til å håndtere kortsiktige forpliktelser ved å selge eiendeler. En lav verdi av TCE indikerer derimot at bedriften har mindre egenkapital for å tåle tap, og dermed større sannsynlighet for å gå konkurs. Sentralt i økonomisk teori er valg av kapitalstruktur, hvor en stor egenkapitalandel anses som et sunnhetstegn knyttet til hvor solid bedriften er. Dette støttes blant annet av Hillier mfl. (2011), hvor det argumenteres for at en høyere egenkapitalandel medfører mindre risiko, og mer robusthet knyttet til å håndtere uforutsette tap. Det vil derfor være naturlig at en lav egenkapitalandel, som i denne variabelen, tenderer til å trekke konkurssannsynligheten opp.

Variabelen $\ll(\text{kortsiktig gjeld} - \text{kortsiktig likviditet})/\text{totale eiendeler}\gg$ sier noe om en bedrifts evne til å betale sine kortsiktige forpliktelser med sine tilgjengelige midler (kortsiktig likviditet), i forhold til bedriftens totale eiendeler. Beeswarmplottene viser at høye verdier av variabelen tenderer til å øke sannsynligheten for konkurs. Dette er i tråd med økonomisk teori hvor kortsiktig gjeld anses som forpliktelser som må tilbakebetales innen kort tid, samtidig som bedriftens eiendeler vanligvis tar lengre tid å omsette. En høy andel kortsiktig gjeld kan derfor føre til at bedriften ikke evner å møte sine kortsiktige forpliktelser, og dermed havner i økonomisk nød (Dirman, 2020). Her er også likviditeten i bedriften sentralt, på samme måte som for variabelen knyttet til skyldige skatter og avgifter over. Funnene fra modellen er dermed i tråd med økonomisk teori.

For variablene «leverandørgjeld/totalte eiendeler» og «leverandørgjeld/salg» gjelder den samme tilnærmingen knyttet til kortsiktig gjeld. Forholdstallene representerer en bedrifts andel leverandørgjeld, i forhold til totale eiendeler og salg. Beeswarmplottene indikerer at en høy verdi tenderer til å øke sannsynligheten for konkurs for begge forholdstallene. Motsatt vil en lav verdi redusere konkurssannsynligheten. Det bør likevel nevnes at variabelen «leverandørgjeld/salg» også har enkelte lave verdier i flere folder og har en tendens til å øke konkurssannsynligheten. Dette gjelder dog ikke for majoriteten av observasjonene. En generell oppfatning i økonomisk teori er at en høy kortsiktig gjeldsgrad indikerer at bedriften er mindre robust, og dermed en faktor som øker sannsynligheten for konkurs (Dirman, 2020). Funnene fra modellene er derfor i stor grad i tråd med økonomisk teori, og anses som relativt pålitelige.

Variabelen «renteinntekter/renteutgifter» representerer en bedrifts evne til å betjene sin gjeld, og er også på mange måter et rentabilitetsmål for hvor godt bedriften evner å generere overskudd. Beeswarmplottene viser at lave verdier av variabelen tenderer til å øke konkurssannsynligheten, og at høye verdier reduserer konkurssannsynligheten. Når renteinntektene er høyere enn renteutgiftene, evner bedriften å generere overskudd og økt lønnsomhet, og dermed redusert konkurssannsynlighet. Tendensene som modellen fanger opp kan derfor anses som pålitelige.

Variabelen «utbytte/årsresultat» er generelt en viktig variabel med høy standardisert SHAP-verdi i testårene 2010-2014, og videre redusert viktighet i resterende år. Modellene fra 2010 til 2012 viser klare tendenser til at en lav verdi av variabelen, altså lavere utbyttebetalinger, øker konkurssannsynligheten. Utbytteutbetalinger brukes vanligvis som et virkemiddel av bedriftens ledelse for å signalisere at bedriften har god økonomisk helse og dermed reduserer konkurssannsynligheten (Frankfurter mfl., 2003). Videre finner Ouma mfl. (2012) en positiv korrelasjon mellom utbytteutbetalinger og bedriftens økonomiske situasjon, og indikerer dermed at en høy verdi av variabelen ville vært et sunnhetstegn. Derimot presenterer Frankfurter mfl. (2003) at utbytteutbetalinger fører til reduserte midler tilgjengelig for å møte sine forpliktelser, som naturlig vil øke konkurssannsynligheten. Til tross for at de første tre testårene viser klare tendenser knyttet til lave verdier, ser vi enkelte avvik i resterende testår. Disse viser likevel en klar konsentrasjon av lave verdier, med enkelte enkelte firmaår med høye variabelverdier som bidrar til å trekke SHAP-verdien opp. Dette kan være firmaår med variabelverdier som er høyere enn majoriteten. Det kan derfor se ut til at modellene fanger opp det positive signalet knyttet til utbytteutbetalinger, samtidig som modellene fanger opp den negative likviditetseffekten av høye utbytteutbetalinger i enkelte firmaår. Totalt sett ser hovedtendensen av modellen

til å være pålitelig.

For variabelen « $\log(\text{totale eiendeler})$ » finner modellen at lave verdier øker konkurs-sannsynligheten. Generelt sett vil høyere totale eiendeler gi bedre økonomisk sikkerhet, og dermed lavere konkurssannsynlighet. Økonomisk teori presenterer at en bedrift med høye totale eiendeler vil stå godt rustet til å håndtere eventuelle fremtidig krevende betalingsforpliktelser enn bedrifter med lave totale eiendeler. Funnene i modellen er derfor i tråd med økonomisk teori og sunn fornuft.

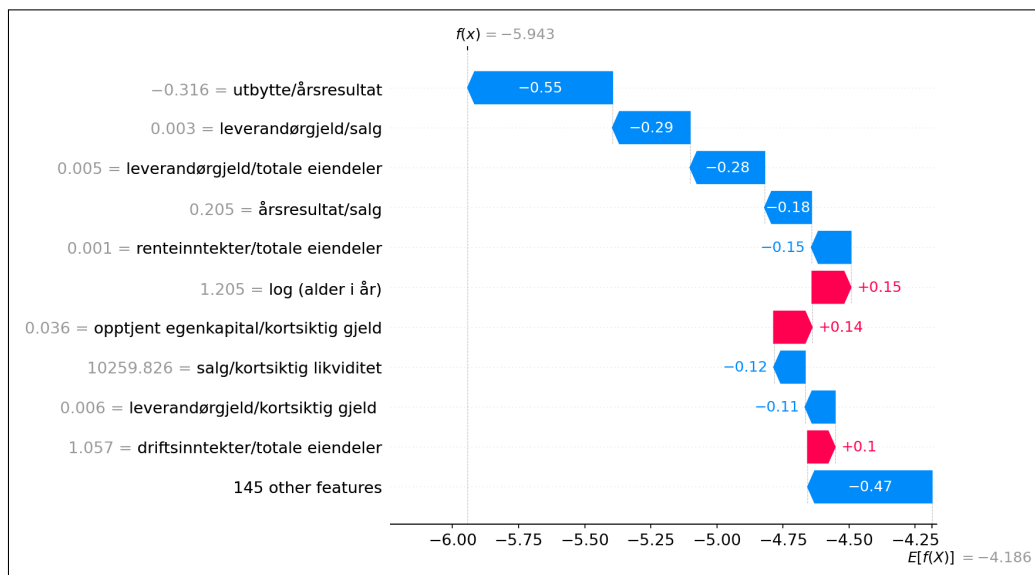
Variabelen «opptjent egenkapital/kortsiktig gjeld» gir oss forholdet mellom opptjent egenkapital og kortsiktig gjeld. Den opptjente egenkapitalen representerer egenkapitalen som over tid er spart opp og ikke utbetalt til aksjonærene. En høy verdi av denne posten indikerer bedre økonomisk sikkerhet. Videre indikerer som nevnt en høy andel kortsiktig gjeld at selskapet er mindre robust. Samlet sett ut ifra økonomisk teori vil derfor en lav verdi av «opptjent egenkapital/kortsiktig gjeld» tyde på svakere økonomi. Ut i fra beeswarmplottene ser man at høye verdier tenderer til å redusere sannsynligheten for konkurs.

SHAP beeswarmplott kan med andre ord med fordel brukes for å analysere en modells globale variabelviktighet, hvor modellen blir mer tolkbar ved å få innblikk i hvilke variabler som har størst betydning for modellen. Alt i alt finner vi at variableffektene som modellen har fanget opp, som fremkommer i beeswarmplottene, i stor grad er i tråd med økonomisk teori, tidligere litteratur og intuisjon. Modellen evner med andre ord å fange opp generelle tendenser fra den virkelige verden, som gir grunnlag for å si at funnene fra modellen er pålitelige. SHAP-rammeverket evner derfor å øke tolkbarheten til XGBoost-modellene ved forklaring av den globale variabelviktigheten.

5.3.2 Lokal forklaring

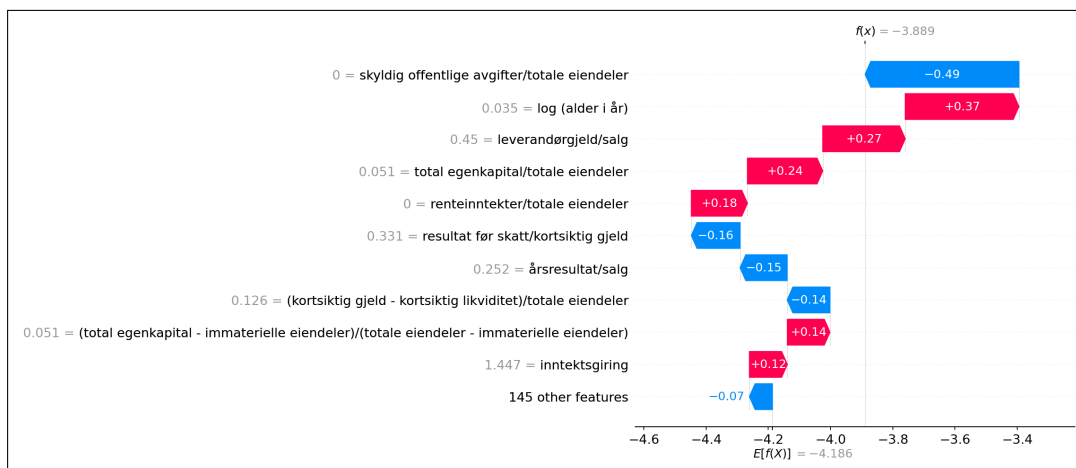
For å enklere kunne tolke hvordan modellen foretar enkeltbeslutninger, kan vi finne ut hvilke variabler som påvirker modellen til å klassifisere den ene klassen fremfor den andre. Vi har i den anledning valgt å bruke SHAP *waterfall-plot* for å illustrere disse. Vi har plukket ut fire firmaår som skal undersøkes: (1) en konkurs med lav sannsynlighet for konkurs, (2) en konkurs med høy sannsynlighet for konkurs, (3) en ikke konkurs med lav sannsynlighet for konkurs og (4) en ikke konkurs med høy sannsynlighet for konkurs. Observasjonene er hentet fra fold 2010. Nederst i waterfall plottene presenteres den gjennomsnittlige predikerte log oddsen for modellen lik

-4,186. Dette betyr at den predikerte sannsynligheten for konkurs av modellen på tvers av alle observasjonene i treningssettet er $\frac{e^{-4,186}}{1+e^{-4,186}} = 1,50\%$. Vi har for enkelhets skyld valgt å presentere kun de ti viktigste variablene for hver av de fire observasjonene. I tillegg vises bidraget fra de resterende 145 variablene i en samlest post nederst.



Figur 3: Lokal forklaring. Konkursbedrift med lav sannsynlighet for konkurs.

I figur 3 presenteres et firmaår som er konkurs ($y=1$), men som har lav sannsynlighet for konkurs. Av de 10 variablene, er det kun tre variabler som bidrar til økt sannsynlighet for konkurs. Derimot ser vi at de fem viktigste variablene, «utbytte/årsresultat», «leverandørgjeld/salg», «leverandørgjeld/totalte eiendeler», «årsresultat/salg» og «renteinntekter/totalte eiendeler», reduserer sannsynligheten for konkurs. Totalt sett reduseres konkurssannsynlighet for firmaåret i forhold til gjennomsnittet i modellen, med en predikert log odds lik -5,943. Dette representerer en sannsynlighet for konkurs lik $\frac{e^{-5,943}}{1+e^{-5,943}} = 0,26\%$.



Figur 4: Lokal forklaring. Konkursbedrift med høy sannsynlighet for konkurs.

I figur 4 presenteres et firmaår som er konkurs ($y=1$), og har høy sannsynlighet for konkurs. Totalt sett trekker seks av de ti viktigste variablene prediksjonen i mot økt sannsynlighet for konkurs. Derimot trekker den viktigste variabelen «skyldig offentlige avgifter/totale eiendeler», mot reduserte sannsynlighet for konkurs. Dette er som forventet, siden variabelen har verdi 0, altså ingen skyldig offentlige avgifter. På den andre siden trekker de neste fire variablene, «log(alder i år)», «leverandørgjeld/salg», «total egenkapital/totale eiendeler» og «renteinntekter/totale eiendeler», prediksjonen mot økt sannsynlighet for konkurs. Dette trekker totalt sett firmaåret mot høyere konkurssannsynlighet med en predikert log odds lik -3,889. Dette representerer en sannsynlighet for konkurs lik $\frac{e^{-3,889}}{1+e^{-3,889}} = 2,01\%$.

Tabell 6: Forventet verdi for variablene i figur 4.

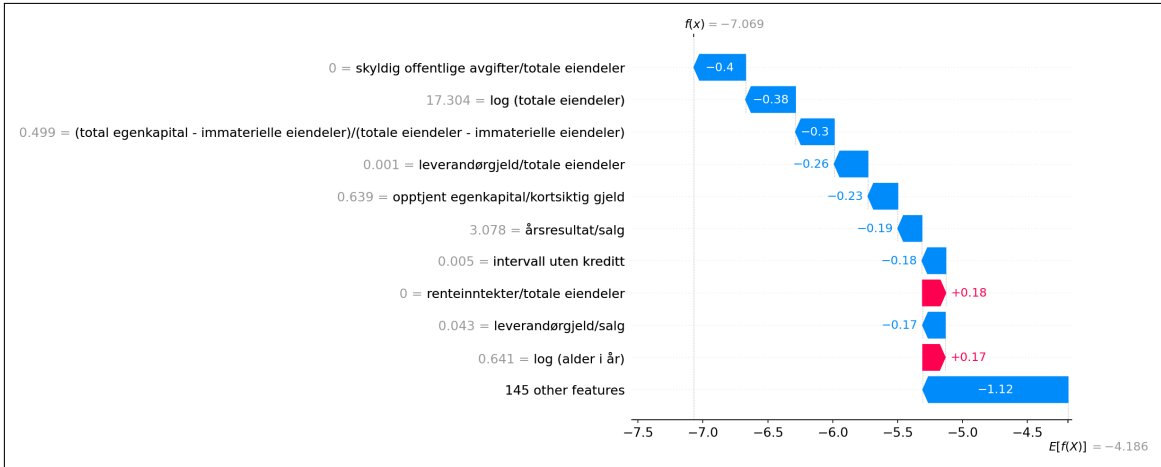
Variabler	Forventet verdi
Skyldig offentlige avgifter/totale eiendeler	0,073
Log (alder i år)	1,907
Leverandørgjeld/salg	0,047
Total egenkapital/totale eiendeler	0,292
Renteinntekter/totale eiendeler	0,005
Resultat før skatt/kortsiktig gjeld	0,192
Årsresultat/salg	0,057
(kortsiktig gjeld – kortsiktig likviditet)/totale eiendeler	0,373
(tot. EK – imm. eiendeler)/(tot. eiendeler – imm. eiendeler)	0,448
Inntektsgiring	- 0,040

I tabell 6 presenteres gjennomsnittlig forventet verdi i modellen for variablene med høyest viktighet for beslutningen i figur 4. Ved å sammenligne variabelverdiene som ligger til grunn for beslutningen med de gjennomsnittlige verdiene for variabelen i

modellen, kan vi undersøke hvorvidt den lokale forklaringen er i tråd med den globale forklaringen av modellen, som presentert i figur 17. Dette gjøres ved å undersøke om variabelverdien i beslutningen er relativt høyere eller lavere enn gjennomsnittet, og dermed vurdere om variabeleffektene er konsistente med funnene i kapittel 5.3.1.

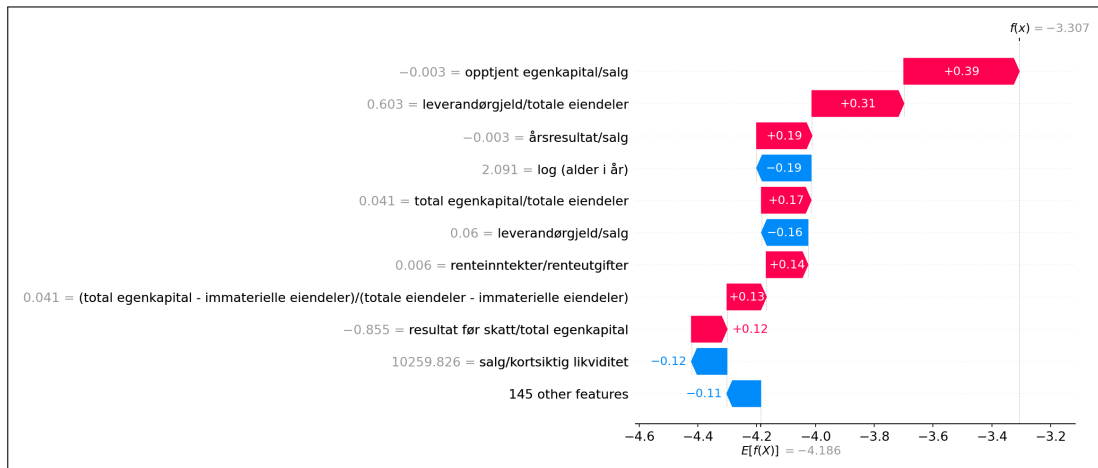
For den viktigste variabelen, «skyldig offentlige avgifter/totale eiendeler», ser vi at variabelverdien er 0, og dermed lavere enn den gjennomsnittlige variabelverdien. I tråd med den globale forklaringen av modellen, så ser det ut som modellen fanger opp den positive likviditetseffekten ved at variabelverdien er lav, og reduserer den predikerte konkurssannsynligheten basert på denne variabelverdien. Dette er i tråd med funn i den globale forklaringen av modellen, som indikerer at en lav verdi av denne variabelen vil trekke konkurssannsynligheten ned. Det samme gjelder for variabelen «log (alder i år)» hvor observasjonens variabelverdi er betydelig lavere enn den gjennomsnittlige variabelverdien for modellen. Denne variabelen øker konkurssannsynligheten i størst grad, og er dermed konsistent med utledningen av den negative korrelasjonen mellom bedrifts alder og konkurssannsynlighet i den globale forklaringen av modellen. Dette kan sies om samtlige variabler, hvor effektene for dette firmaåret er i tråd med effektene som nevnes i kapittel 5.3.1.

Om bedriften som representeres av dette firmaåret kommer til banken med denne SHAP-forklaringen og lurer på hvorfor banken valgte å avslå en lånesøknad, ville banken kunne forklart i korte trekk hvilke faktorer som har lagt grunnlaget for beslutningen. Banken vil her være i stand til å forklare at bedriften er av ung alder, i tillegg til å ha en stor andel leverandørgjeld på hele 45% som øker konkurssannsynligheten i størst grad. Videre har bedriften en egenkapitalandel på 5%, som tyder på at bedriften er lite robust. Alderen er det ikke noe bedriften kan gjøre noe med, men banken kan her komme med en oppfordring om å balansere kapitalstrukturen mot mer langsiktig finansiering i form av større andel egenkapital, og også langsiktig gjeld. Det bør også nevnes at den variabelen som er av størst viktighet for modellen, «skyldig offentlige avgifter/totale eiendeler», reduserer konkurssannsynligheten, og at dette er et godt tegn. For banken vil formålet med en slik tilbakemelding være at bedriften skal få innsikt i hvilke faktorer de må forbedre for å kunne komme tilbake på et senere tidspunkt for å for innvilget lån, ved å redusere konkurssannsynligheten. For banken vil dette kunne medføre at de ikke nødvendigvis taper kunden, men heller bidrar med en form for rådgivning for hva som skal til for at kunden skal få innvilget lånet, og forblir kunde hos banken.



Figur 5: Lokal forklaring. Ikke-konkursbedrift med lav sannsynlighet for konkurs.

I figur 5 presenteres et firmaår som er ikke konkurs ($y=0$), med lav sannsynlighet for konkurs. Totalt sett ser vi her at åtte av de ti viktigste modellene trekker prediksjonen mot ikke konkurs. Det er kun to variabler, «renteinntekter/total inntekter» og «log(alder i år)», som reduserer sannsynligheten for konkurs. Den viktigste variabelen for prediksjonen av dette firmaåret er «skyldig offentlige avgifter/total eiendeler» med høyeste absolutte SHAP-verdi på $-0,4$. Totalt sett reduseres konkurssannsynligheten i forhold til modellens gjennomsnitt, med en predikert log odds på $-7,069$. Dette representerer en sannsynlighet for konkurs lik $\frac{e^{-7,069}}{1+e^{-7,069}} = 0,09\%$.



Figur 6: Lokal forklaring. Ikke-konkursbedrift med høy sannsynlighet for konkurs.

I figur 6 presenteres et firmaår som er ikke konkurs ($y=0$), men med høy sannsynlighet for konkurs. Her ser vi at de tre viktigste variablene, «opptjent egenkapital/salg», «leverandørgjeld/total eiendeler» og «årsresultat/salg», øker sann-

synligheten for konkurs. I tillegg er det ytterligere fire variabler som øker konkurranssynligheten. Videre er «log(alder i år)» den fjerde viktigste variabelen, og denne reduserer konkurranssynligheten. Totalt sett øker konkurranssynligheten for dette firmaåret, med en predikert log odds på -3,305. Dette representerer en sannsynlighet for konkurs på $\frac{e^{-3,307}}{1+e^{-3,307}} = 3,53\%$.

Tabell 7: Forventet verdi for variablene i figur 6.

Variabler	Forventet verdi
Opptjent egenkapital/salg	0,092
Leverandørgjeld/total eieendeler	0,127
Årsresultat/salg	0,057
Log (alder i år)	1,907
Total egenkapital/total eieendeler	0,292
Leverandørgjeld/salg	0,047
Renteinntekter/renteutgifter	0,178
(tot. EK – imm. eieendeler)/(tot. eieendeler – imm. eieendeler)	0,448
Resultat før skatt/total egenkapital	0,487
Salg/kortsiktig likviditet	68,927

I tabell 7 presenteres gjennomsnittlig forventet verdi i modellen for variablene som presenteres med høyest viktighet for beslutningen presentert i figur 6. Den viktigste variabelen for beslutningen er «opptjent egenkapital/salg», hvor firmaårets variabelverdi er mye lavere enn den gjennomsnittlige variabelverdien for variabelen i modellen. Denne variabelen undersøkes ikke i den globale forklaringen av modellen, da den har relativt lav SHAP-verdi på tvers av foldene. I kapittel 5.3.1 omtales derimot en høy verdi av opptjent egenkapital som et positivt tegn på økonomisk sikkerhet, og en det er dermed naturlig at en negativ verdi av opptjent egenkapital øker konkurranssynligheten for dette firmaåret. Videre er «leverandørgjeld/total eieendeler» av høy variabelviktighet for firmaåret, hvor variabelverdien er betydelig høyere enn den gjennomsnittlige verdien for variabelen i modellen. Det er tydelig at modellen har fanget opp den negative effekten som er knyttet til en stor andel kortsiktig gjeld, og dermed øker konkurranssynligheten. For firmaåret ser vi videre at bedriftens alder er høyere enn gjennomsnittet, og at dette reduserer konkurranssynligheten. Avslutningsvis ser vi at effektene som ble adressert av den globale forklaringen i kapittel 5.3.1, også er konsistent for resterende variabler i dette firmaåret.

Med en så høy sannsynlighet for konkurs, er dette en bedrift som banken sannsynligvis ikke ville innvilget lån til. Om bedriften ber om en forklaring på grunnlaget for beslutningen, kan banken derimot bruke waterfallplottet for å forklare hvilke faktorer som har påvirket beslutningen. Først og fremst kan det forklares at de to faktorene

med størst viktighet er knyttet til bedriftens kapitalstruktur. Forholdet mellom opp-tjent egenkapital og salgsinntekter er negativt, som kan indikere at bedriften tapt mer penger enn de har tjent i forhold til salgsinntektene. Videre ses en svært høy andel av leverandørgjeld i kapitalstrukturen på hele 60,3%, og som nevnt ved flere anledninger i kapittel 5.3.1, vil en stor andel kortsiktig gjeld gjøre bedriften mindre robust. Videre indikerer variabelen «årsresultat/salg», at bedriften går med netto tap etter at driftskostnader og skatter er fratrukket de opprinnelige salgsinntektene. Bedriften oppfattes dermed som uskikket til å generere tilstrekkelig med inntekter for å opprettholde driften. Derimot trekker variabelen «leverandørgjeld/salg» bedriften mot ikke konkurs, hvor leverandørgjelden utgjør kun 6% av bedriftens salgsinntekter. Det tyder derfor på at salgsinntektene til bedriften er relativt høye, men at kostnadene er enda høyere slik at bedriften går av med netto tap. Banken bør i denne situasjonen argumentere for at en ubalansert kapitalstruktur i kombinasjon med tilsynelatende lav evne til å generere overskudd, øker konkurssannsynligheten i stor grad. Banken bør her presentere eventuelle løsninger for disse utfordringene, hvor et sentralt aspekt er at bedriften bør fokusere på økt effektivitet og strengere kostnadsstyring. Dessuten bør det det være økt fokus på å balansere kapitalstrukturen mot mer langsiktig finansiering i form av økt egenkapital i første rekke, og deretter langsiktig gjeld.

Resultatene fra de lokale forklaringene, viser at waterfall plott kan brukes for å tolke enkeltbeslutninger, som vi omtaler som svarte bokser i kapittel 2.4. Generelt observerer vi at variabeleffekten er relativt konsistent mellom de fire firmaårene. Spesielt ser vi at modellen klarer å fange opp effekten av variabelen « $\log(\text{alder i år})$ », da en lav alder øker konkurssannsynligheten i figur 4, og en høy alder reduserer konkurssannsynligheten i figur 6. Til tross for at vi undersøker to ulike firmaår med høy sannsynlighet for konkurs, så ser vi forskjeller i variabelviktighet, hvor den viktigste variabelen i hver av firmaårene, ikke er innenfor de ti viktigste variablene for den andre. Dette indikerer at XGBoost-modellen evner å gjennomføre individuelle vurderinger av de ulike variabelverdiene for en bestemt bedrift. Deretter finner vi at ved å analysere effektene av de ulike variabelverdiene i waterfallplottene, så kan vi vurdere hvorvidt funnene for enkeltobservasjonene er i tråd med de generelle tendensene som XGBoost-modellen finner. Generelt for alle firmaårene ser vi at variabeleffektene er konsistente med de generelle tendensene i XGBoost-modellen, som antyder at XGBoost-modellen gir troverdige prediksjoner. Avslutningsvis kan det argumenteres for at bedrifter med høy sannsynlighet for konkurs burde ta utgangspunkt i den lokale forklaringen for bedrifter med lav sannsynlighet for konkurs for å se hvilke variabeleffekter som reduserer konkurssannsynligheten i størst grad.

SHAP forklarer hvilke variabler som har predikert sannsynlighet for konkurs i den ene eller andre retningen. Dette gir først og fremst bedrifter innblikk i hvilke variabler som har påvirket prediksjonen i mest, og hvilke beslutninger som bør tas for å redusere konkurssannsynligheten. Med disse plottene blir det derfor i større grad transparens i beslutningen. Videre gir det banker innblikk i hvilke variabler som påvirker konkurssannsynligheten for en gitt bedrift. For eksempel i begrunnelsen for hvorvidt de velger å utstede lån, samt at man følger GDPRs krav om transparens og kundenes rett på innsyn. Dermed muliggjør SHAP at banken kan gå tilbake til den enkelte beslutning som er tatt av modellen for å finne forklaring på hvordan beslutningen ble konkludert som den gjorde.

5.4 Anvendelse i praksis

Som nevnt i kapittel 2.1.1 kreves det at prediksjonsmodellen skal være både nøyaktig og tolkbar for at konkursprediksjonen anses som effektiv. Derimot ble det antydning at forholdet tolkning og ytelse må balanseres, hvor ulike brukere har ulike preferanser og fokus på hver av disse. En bank ble argumentert for å ha spesielt fokus på modellens nøyaktighet, hvorimot bedrifter og deres interessenter gjerne har fokus på modellens tolkbarhet. Modellens nøyaktighet er av stor viktighet da dette gir bedre informasjonsgrunnlag for mer lønnsom beslutningstaking. Derfor vil mer kompliserte maskinlæringsmodeller, som generelt sett viser seg å ha bedre prediksjonsevne og nøyaktighet, være av stor interesse og gevinst for både banker og bedrifter.

Videre presenteres en sentral begrensning for anvendelse av komplekse maskinlæringsmodeller i kapittel 2.1, som at den norske finanssektoren er tilknyttet streng regulering og lovgivning. Reguleringen krever høy grad av transparens og ansvarlighet, hvor grunnlaget for enkeltbeslutninger skal kunne begrunnes (EUR-Lex, 2016). Ved å ta i bruk SHAP-rammeverket for tolkning av XGBoost-modellen har vi oppnådd innsikt i modellenes generelle logikk og adferd, og dermed gjort modellen mer tolkbar. Dette har vi oppnådd både på globalt nivå ved å se på XGBoost-modellens lærte tendenser og variabeleffekter, samt på lokalt nivå hvor SHAP-rammeverket har muliggjort innsikt i bakgrunnen for individuelle beslutninger. For banker vil innsikt i modellenes generelle logikk føre til bredere forståelse av hvordan modellene fungerer og hvilke variabler som påvirker prediksjonene. Gjennom grundig utforskning av variabeleffektene oppnås innsikt i graden av pålitelighet og overensstemmelse med økonomisk teori og rasjonell tenkning i prediksjonen. Ved å benytte SHAP-rammeverket forbedres dermed tolkbarheten og påliteligheten til

XGBoost-modellene.

For bedrifter vil tolkbarheten av en konkursprediksjonsmodell være nyttig i form av at modellen kan tas i bruk som et strategisk verktøy knyttet opp mot bedrifters forretningsmodell. Modellen gir beslutningstakere i bedriften innsikt i hvilke utfordringer som øker bedriftens konkurssansynlighet og økonomisk nød. Dessuten gjør den dem også i stand til å ta mer velbegrunnede og informerte beslutninger for å motvirke de negative tendensene som øker konkurssansynligheten. Ved å se på variabelviktighet i modellen er det mulig for bedriftenes beslutningstakere å undersøke hvorvidt de ulike variablene påvirker denne konkurssansynligheten, og hvordan bedriften ligger an i forhold til tendensene i modellen. Dette gjør beslutningstakerne i stand til å kunne gjennomføre justeringer knyttet til sine svake sider, samt styrke sine eksisterende kapabiliteter. De lokale forklaringene gjør det derfor mulig å undersøke den gjeldende bedriften, og se på hvilke spesifikke variabler som eventuelt bør justeres for å potensielt redusere konkurssansynligheten for bedriften.

De lokale forklaringene er også til interesse både for banker og bedrifter knyttet til långivning. Hvis en bedrift mottar avslag på lånesøknad fra banken, og ber om begrunnelse, kan banken bruke de lokale forklaringene som forklaring på avslaget. For bankene vil bruk av lokale forklaringer legge til rette for rådgivning av hvilke variabler som bør justeres for å redusere konkurssansynligheten, og dermed øke bedriftens mulighet for å få innvilget lån på et senere tidspunkt. Dette vil potensielt kunne bidra til å øke tillit- og lojalitetsforholdet mellom banken og bedriften, og gir kunden økt insentiv til å forbli kunde hos banken. Derimot presenterer Slack mfl. (2020) at det er en stor utfordring knyttet til anvendelse av SHAP som tolkningsteknikk for komplekse modeller, fordi det er enkelt å produsere misvisende tolkninger. Det er derfor viktig at banker utøver stor grad av transparens knyttet til hvordan de gjennomfører SHAP-forklaringer.

I kapittel 2.1.1 presenterer Bracke mfl. (2019) tre sentrale spørsmål knyttet til tolkning av maskinlæringsmodeller. Det første spørsmålet er knyttet til hva som driver forklaringene generelt, og de globale forklaringene som SHAP-rammeverket presenterer evner å gi innblikk i modellens generelle lærte logikk, samt hvilke variabler som er av størst viktighet for konkursprediksjonsmodellen. Videre er neste spørsmål knyttet til hvilke variabler og hvilke variabeleffekter som påvirker de individuelle prediksjonene. I denne studien ser vi at SHAP-rammeverket evner og legger til rette for å gjennomføre lokale forklaringer, og dermed øke tolkbarheten uten å redusere modellens nøyaktighet og prediksjonsevne. Det tredje spørsmålet er knyttet til hvor-

dan modellen fungerer, og hvorvidt dette enkelt kan forklares. Her er det fortsatt utfordringer, da SHAP-rammeverket er bygd opp av avansert matematikk, prosedyrer og algoritmer. Uansett, ser vi at både globale og lokale forklaringer med SHAP øker XGBoost-modellenes tolkbarhet. Dette besvarer derfor med andre ord studiens forskningsspørsmål som definert i kapittel 1.2.

6 Konklusjon

I denne masteroppgaven undersøker vi hvordan prediksjoner gjort av komplekse maskinlæringsmodeller kan bedres ved anvendelse av SHAP-rammeverket innenfor konkursprediksjon. I dette kapittelet oppsummerer vi kort de mest sentrale funnene ved studien. Avslutningsvis presenterer vi mulige svakheter ved studien, samt muligheter for videre forskning på området.

I denne studien forsøker vi å undersøke følgende forskningsspørsmål: «*Hvordan kan xAI øke tolkbarheten av komplekse maskinlæringsmodeller for konkursprediksjon, og hvor aktuelt er det å ta i bruk dette i finanssektoren?*». Vi starter derfor med å undersøke om XGBoost har bedre prediksjonsevne enn logistisk regresjon på våre data, ved å analysere modellenes prediksjonsevne ved bruk av AUC og Brier-score. Vi finner at XGBoost har bedre prediksjonsevne enn logistisk regresjon i samtlige folder, med en gjennomsnittlig AUC på 90% out-of-sample, mot logistisk regresjon med gjennomsnittlig AUC på 87%. I tillegg har XGBoost gjennomgående lavere Brier Score. XGBoost-modellene har med andre ord bedre prediksjonsevne.

Logistisk regresjon benyttes fremfor mer fleksible maskinlæringsmetoder som XGBoost fordi prediksjonen ved logistisk regresjon ofte er mer tolkbare. Med utgangspunkt i den økte prediksjonsevnen som følger av anvendelsen av XGBoost, undersøker vi hvorvidt SHAP-rammeverket evner å øke tolkbarheten til XGBoost-modellene. Først analyserer vi bruken av SHAP for globale forklaringer, hvor vi finner at modellenes generelle logikk og lærte adferd i stor grad er i tråd med både økonomisk teori og sunn fornuft. Dette øker følgelig graden av pålitelighet for XGBoost-modellene. Deretter analyserer vi bruk av SHAP for lokale forklaringer, hvor enkeltbeslutninger i modellen evalueres. Her finner vi at SHAP gjør det mulig å tolke hvilke variabler som er av størst påvirkning for beslutningen, samt hvilken effekt variablene har på prediksjonen. Vi finner derfor med andre ord at anvendelse av SHAP gjøre det mulig å tolke både på globalt og lokalt nivå, samtidig som at modellens prediksjonsevne opprettholdes. Resultatene i denne studien indikerer at SHAP-rammeverket kan være en effektiv løsning på utfordringen knyttet til svarte bokser, og dermed gjøre komplekse maskinlæringsmodeller mer anvendbare for konkursprediksjon i finanssektoren.

6.1 Svakheter ved studien

I denne studien predikerer vi konkurs ett år frem i tid. I den virkelige verden vil dette være nyttig, men ikke ideelt da banker som utsteder lån i stor grad ønsker å kunne forutse sannsynlighet for konkurs og mislighold av gjeld for en lengre periode enn et år frem i tid. I tillegg vil en prediksjonshorisont på ett år ha begrenset tilgang på relevant informasjon, hvor bedrifter for eksempel gjennomfører strategiske beslutninger som gir midlertidige tilsynelatende negative tendenser i regnskapet, som har til hensikt å generere bedre resultater på lengre sikt. Med en såpass kort prediksjonshorisont er det derfor en risiko for at modellen ikke klarer å fange opp alle fremtidige konkurstendenser. Samtidig kan modellen fange opp midlertidige konkurstendenser som i teorien ikke nødvendigvis utgjør økt konkurssannsynlighet.

I denne studien tar vi kun i bruk informasjon fra årsregnskaper for å predikere konkurs. Regnskapsføring er strengt regulert, og anses som i stor grad standardisert. I studien antar vi at årsregnskapene er rapportert på riktig måte. Dette er ikke dermed sagt en 100% realistisk antakelse i den virkelige verden, men på den andre siden likevel et godt utgangspunkt for denne studien.

For å bevare samme antall firmaår og samtidig forsøke å forbedre prediksjonsnøyaktigheten valgte vi å gjennomføre winsorizing på 1. og 99. persentil. Det betyr at enkelte ekstremverdier blir satt til verdien innenfor persentilene og på den måten kan være med på å påvirke resultatet. I den virkelige verden kan ikke slike forutsetninger tas på samme måte som i denne studien da det eksempelvis kan være en bedrift som befinner seg utenfor 1. og 99. persentil. Dette kan redusere påliteligheten til den lokale og globale SHAP forklaringen av enkelte prediksjoner.

SHAP-rammeverket virker å gi relativt konsistente forklaringer i vår XGBoost-modell, og vi vurderer derfor i kapittel 5.4 hvorvidt en kompleks modell som tolkes av SHAP kan anvendes i praksis i den virkelige verden. Basert på Fryer mfl. (2021) og Balcaen og Ooghe (2006), finner vi at SHAP ikke nødvendigvis egner seg for variabelseleksjon, og at det er klare forskjeller mellom de viktigste variablene for konkurs prediksjon mellom ulike datautvalg og land. Dette indikerer derfor at til tross for at resultatene i studien er i tråd med økonomisk teori og litteratur, så vil det være naturlig å antyde at de tendensene og variabeleffektene som kommer frem av denne studien i liten grad kan generaliseres til andre datautvalg enn norske SMB.

6.2 Videre forskning

Studien vår viser at XGBoost har bedre prediksjonsevne enn logistisk regresjon, og at XGBoost-modellen fint kan tolkes ved hjelp av SHAP. Et sentralt spørsmål som da stilles er hvorfor XGBoost ikke tas i bruk av bankene. Det er derfor relevant å gå grundigere inn på denne utfordringen og finne ut hva som skal til for at XGBoost med SHAP faktisk kan tas i bruk i finanssektoren, og hvilke tilpasninger som kreves.

Videre velger vi i studien å se på konkursprediksjon på tvers av industrier, uten å ta hensyn til ulikheter innad i disse. Det er naturlig å tenke at det er forskjellige variabler som skiller seg ut i de ulike industriene. Ved å se på en industri om gangen kan man lage modeller som er bedre tilpasset den spesifikke industrien. På den måten er det muligheter for at prediksjonsevnen til modellene forbedres, og det kan derfor være enda et insentiv til å ta i bruk XGBoost med SHAP.

I kapittel 6.1 presenterer vi utfordringer med at en prediksjonshorisont på ett år kan anses som kort. Det er derfor relevant å gjennomføre studier med lengre prediksjonshorisont, for å møte bankers interesse om å kunne forutsi konkurssannsynligheten lengre frem i tid. Det kan videre tenkes at tidsperiodene for de valgte foldene og treningssettene har noe å si for resultatene, da den økonomiske situasjonen til bedrifter kan bli påvirket ulikt på kort og lang sikt. Et forslag kan derfor være å gjennomføre estimeringen av modellene på nytt, men da justere tidsperiodene for trenings- og testsettene. På den måten kan potensielle kortsiktige eller langsiktige trender fanges opp.

Vår studie brukes data som ikke inneholder firmaår fra regnskapsårene som påvirkes av Covid-19 pandemien. En interessant ide kan derfor være å lage modeller som baserer seg på data fra disse årene. Da kan man sammenligne funnene med våre og se om det er noe som skiller seg ut for denne perioden for norske SMB. Det er videre interessant å se om det er andre variabler som skiller seg ut som viktige for konkursprediksjon i disse årene.

Avslutningsvis er det interessant å studere andre komplekse maskinlæringsmetoder, og undersøke hvorvidt SHAP er like effektiv for å øke tolkbarheten av disse. Det er blant annet flere ensemble metoder for trebaserte modeller som har lav iboende tolkbarhet, og dermed anses som svarte bokser. I tillegg ser vi i figur 2 at blant annet nevralt nettverk og støttevektormaskiner er andre maskinlæringsmetoder med relativt lav iboende tolkbarhet, hvor SHAP kan brukes som tolkningsmetode. Det vil derfor etter hvert være interessant å se hvilken maskinlæringsmetode som gir best

prediksjonsevne for konkursprediksjon på norske SMB, samtidig som SHAP evner å øke tolkbarheten for disse metodene.

Bibliografi

- Adadi, A., Berrada, M., 2018. Peeking inside the blackbbbox: A survey on explainable artificial intelligence (xai). *IEEE Access* 6, 52138–52160. doi:10.1109/ACCESS.2018.2870052.
- Agarwal, V., Taffler, R., 2008. Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking & Finance* 32, 1541–1551. doi:10.1016/j.jbankfin.2007.07.014.
- Agarwal, V., Taffler, R.J., 2007. Twenty-five years of the taffler z-score model: Does it really have predictive ability? *Accounting and Business Research* 37, 285–300. doi:10.1080/00014788.2007.9663313.
- Alaka, H.A., Oyedele, L.O., Owolabi, H.A., Oyedele, A.A., Akinade, O.O., Bilal, M., Ajayi, S.O., 2017. Critical factors for insolvency prediction: towards a theoretical model for the construction industry. *International Journal of Construction Management* 17, 25–49. doi:10.1080/15623599.2016.1166546.
- Altman, E.I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance* 23, 589–609. doi:10.2307/2978933.
- 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, 29–54. doi:10.1016/0378-4266(77)90017-6.
- Altman, E.I., Sabato, G., 2007. Modelling credit risk for smes: Evidence from the us market. *Abacus* 43, 332–357. doi:10.1111/j.1467-6281.2007.00234.x.
- Arrieta, A.B., Diaz-Rodriguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., mfl., 2020. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* 58, 82–115. doi:10.1016/j.inffus.2019.12.012.
- Arthur, W.B., 2009. Complexity and the economy, i: Handbook of Research on Complexity. Edward Elgar Publishing. doi:10.4337/9781781952665.00007.
- Babaei, G., Giudici, P., Raffinetti, E., 2021. Explainable fintech lending. Available at SSRN 3892652 doi:10.2139/ssrn.3892652.

-
- Balcaen, S., Ooghe, H., 2006. 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. *The British Accounting Review* 38, 63–93. doi:10.1016/j.bar.2005.09.001.
- Barry, A., Li, W., Becerra, J.A., Gilabert, P.L., 2021. Comparison of feature selection techniques for power amplifier behavioral modeling and digital predistortion linearization. *Sensors* 21, 5772. doi:10.3390/s21175772.
- Bauer, E., Kohavi, R., 1999. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine learning* 36, 105–139. doi:10.1023/A:1007515423169.
- Beaver, W.H., 1966. Financial ratios as predictors of failure. *Journal of Accounting Research* , 71–111doi:10.2307/2490171.
- Bernhardsen, E., Larsen, K., 2007. Modelling credit risk in the enterprise sector- further development of the sebra model. *Norges Bank. Economic Bulletin* 78, 102.
- BIS, 2011. Basel III: A global regulatory framework for more resilient banks and banking systems - revised version june 2011 .
- Bracke, P., Datta, A., Jung, C., Sen, S., 2019. Machine learning explainability in finance: An application to default risk analysis doi:10.2139/ssrn.3435104.
- Breiman, L., 2017. *Classification and Regression Trees*. Routledge, New York. doi:10.1201/9781315139470.
- Brier, G.W., 1950. Verification of forecasts expressed in terms of probability. *Monthly Weather Review* 78, 1–3.
- Brynjolfsson, E., Hitt, L.M., Kim, H.H., 2011. Strength in numbers: How does data-driven decisionmaking affect firm performance? Available at SSRN 1819486 doi:10.2139/ssrn.1819486.
- Brzezinski, D., Stefanowski, J., 2017. Prequential auc: properties of the area under the roc curve for data streams with concept drift. *Knowledge and Information Systems* 52, 531–562. doi:10.1007/s10115-017-1022-8.
- Burrell, J., 2016. How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big data & Society* 3. doi:10.1177/2053951715622512.

-
- Bussmann, N., Giudici, P., Marinelli, D., Papenbrock, J., 2021. Explainable machine learning in credit risk management. *Computational Economics* 57, 203–216. doi:10.1007/s10614-020-10042-0.
- Campbell, J.Y., Hilscher, J., Szilagyi, J., 2008. In search of distress risk. *The Journal of Finance* 63, 2899–2939. doi:10.1111/j.1540-6261.2008.01416.x.
- Carmona, P., Climent, F., Momparler, A., 2019. Predicting failure in the u.s. banking sector: An extreme gradient boosting approach. *International Review of Economics & Finance* 61, 304–323. doi:10.1016/j.iref.2018.03.008.
- Caruana, R., Niculescu-Mizil, A., 2006. An empirical comparison of supervised learning algorithms, i: Proceedings of the 23rd International Conference on Machine learning, s. 161–168. doi:10.1145/1143844.1143865.
- Castelvecchi, D., 2016. Can we open the black box of ai? *Nature News* 538, 20.
- Chandrashekar, G., Sahin, F., 2014. A survey on feature selection methods. *Computers & Electrical Engineering* 40, 16–28. doi:10.1016/j.compeleceng.2013.11.024.
- Chava, S., Jarrow, R.A., 2004. Bankruptcy prediction with industry effects. *Review of finance* 8, 537–569. doi:10.1093/rof/8.4.537.
- Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system, i: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, s. 785–794. doi:10.1145/2939672.2939785.
- Chollet, F., 2021. *Deep learning with Python*. Manning Publications Co., Shelter Island.
- Chou, T.N., 2019. An explainable hybrid model for bankruptcy prediction based on the decision tree and deep neural network, i: 2019 IEEE 2nd International Conference on Knowledge Innovation and Invention (ICKII), IEEE. s. 122–125. doi:10.1109/ICKII46306.2019.9042639.
- Ciatto, G., Schumacher, M.I., Omicini, A., Calvaresi, D., 2020. Agent-based explanations in ai: Towards an abstract framework, i: Explainable, Transparent Autonomous Agents and Multi-Agent Systems: Second International Workshop, EXTRAAMAS 2020, Auckland, New Zealand, May 9–13, 2020, Revised Selected Papers 2, Springer. s. 3–20.

-
- Deakin, E.B., 1972. A discriminant analysis of predictors of business failure. *Journal of Accounting Research* , 167–179doi:10.2307/2490225.
- Dietterich, T.G., 2000. Ensemble methods in machine learning, i: Multiple Classifier Systems: First International Workshop, MCS 2000 Cagliari, Italy, June 21–23, 2000 Proceedings 1, Springer. s. 1–15.
- Dirman, A., 2020. Financial distress: the impacts of profitability, liquidity, leverage, firm size, and free cash flow. *International Journal of Business, Economics and Law* 22, 17–25.
- Dong, G., Lai, K.K., Yen, J., 2010. Credit scorecard based on logistic regression with random coefficients. *Procedia Computer Science* 1, 2463–2468. doi:10.1016/j.procs.2010.04.278.
- Dovsilovic, F.K., Brvcic, M., Hlupic, N., 2018. Explainable artificial intelligence: A survey, i: 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), IEEE. s. 210–215. doi:10.23919/MIPRO.2018.8400040.
- Du, M., Liu, N., Hu, X., 2019. Techniques for interpretable machine learning. *Communications of the ACM* 63, 68–77. doi:10.1145/3359786.
- Edmister, R.O., 1972. An empirical test of financial ratio analysis for small business failure prediction. *Journal of Financial and Quantitative Analysis* 7, 1477–1493. doi:10.2307/2329929.
- Eklund, T., Larsen, K., Bernhardsen, E., 2001. Modell for analyse av kredittrisiko i foretakssektoren. *Penger og Kreditt* , 109–116<https://norges-bank.brage.unit.no/norges-bank-xmlui/bitstream/handle/11250/2480734/eklund.pdf?sequence=1&isAllowed=y>.
- von Eschenbach, W.J., 2021. Transparency and the black box problem: Why we do not trust ai. *Philosophy & Technology* 34, 1607–1622. doi:10.1007/s13347-021-00477-0.
- EUR-Lex, 2016. Regulation (EU) 2016/679 of the European parliament and the council of 27 april 2016 Tilgjengelig fra: <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02016R0679-20160504> (Hentet: 7. april 2023).

-
- Filipe, S.F., Grammatikos, T., Michala, D., 2016. Forecasting distress in european sme portfolios. *Journal of Banking & Finance* 64, 112–135. doi:10.1016/j.jbankfin.2015.12.007.
- Frankfurter, G., Wood, B.G., Wansley, J., 2003. Dividend policy: Theory and practice. Elsevier, Oxford.
- Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics* , 1189–1232.
- Fryer, D., Strümke, I., Nguyen, H., 2021. Shapley values for feature selection: The good, the bad, and the axioms. *IEEE Access* 9, 144352–144360. doi:10.1109/ACCESS.2021.3119110.
- Hautsch, N., Schaumburg, J., Schienle, M., 2015. Financial network systemic risk contributions. *Review of Finance* 19, 685–738. doi:10.1093/rof/rfu010.
- Hicks, J.R., 1962. Liquidity. *The Economic Journal* 72, 787–802. doi:10.2307/2228351.
- Hillier, D., Grinblatt, M., Titman, S., 2011. *Financial Markets and Corporate Strategy*. 2 utgave, McGraw Hill, Berkshire.
- Hosmer Jr., D.W., Lemeshow, S., Sturdivant, R.X., 2013. *Applied logistic regression*. volum 398. John Wiley & Sons, Hoboken.
- Härdle, W., Lee, Y.J., Schäfer, D., Yeh, Y.R., 2009. Variable selection and over-sampling in the use of smooth support vector machines for predicting the default risk of companies. *Journal of Forecasting* 28, 512–534. doi:10.1002/for.1109.
- James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. *An Introduction to Statistical Learning*. volum 112. Springer, New York. doi:10.1007/978-1-0716-1418-1.
- Jones, S., Johnstone, D., Wilson, R., 2015. An empirical evaluation of the performance of binary classifiers in the prediction of credit ratings changes. *Journal of Banking & Finance* 56, 72–85. doi:10.1016/j.jbankfin.2015.02.006.
- Jones, S., Johnstone, D., Wilson, R., 2017. Predicting corporate bankruptcy: An evaluation of alternative statistical frameworks. *Journal of Business Finance & Accounting* 44, 3–34. doi:10.1111/jbfa.12218.
- Kolari, J.W., Shin, G.H., 2006. Assessing the profitability and riskiness of small business lenders in the banking industry. Available at SSRN 1017093 doi:10.2139/ssrn.1017093.

-
- Konkursloven, 1984. Lov om gjeldsforhandling og konkurs (konkursloven) Tilgjengelig fra: <https://lovdata.no/dokument/NL/lov/1984-06-08-58> (Hentet: 20 april 2023).
- Konkursrådet, 2012. Innføring i konkurs Tilgjengelig fra: <https://konkursradet.custompublish.com/innfoering-i-konkurs.305266.no.html> (Hentet: 24. februar 2023).
- Kotsiantis, S., Kanellopoulos, D., Pintelas, P., mfl., 2006. Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering* 30, 25–36.
- Kotsiantis, S.B., Zaharakis, I., Pintelas, P., mfl., 2007. Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering* 160, 3–24.
- Kumar, P.R., Ravi, V., 2007. Bankruptcy prediction in banks and firms via statistical and intelligent techniques—a review. *European Journal of Operational Research* 180, 1–28. doi:10.1016/j.ejor.2006.08.043.
- Laborda, J., Ryoo, S., 2021. Feature selection in a credit scoring model. *Mathematics* 9, 746. doi:10.3390/math9070746.
- Laitinen, T., Kankaanpaa, M., 1999. Comparative analysis of failure prediction methods: the finnish case. *European Accounting Review* 8, 67–92. doi:10.1080/096381899336159.
- Lee, T.S., Chiu, C.C., Chou, Y.C., Lu, C.J., 2006. Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics & Data Analysis* 50, 1113–1130. doi:10.1016/j.csda.2004.11.006.
- Lervik, F., 2022. Brytningstid - en studie av kunstig intelligens i norske banker og forsikringsselskap Tilgjengelig fra: <https://www.finansforbundet.no/content/uploads/2022/08/KI-i-norske-banker-og-forsikringsselskap-endelig-versjon.pdf> (Hentet: 14. april 2023).
- Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R.P., Tang, J., Liu, H., 2017. Feature selection: A data perspective. *ACM computing surveys (CSUR)* 50, 1–45. doi:10.1145/3136625.

-
- 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, 561–572. doi:10.1016/j.ejor.2016.01.012.
- Lipton, Z.C., 2018. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue* 16, 31–57.
- Lundberg, S., 2018a. Beeswarm plot Tilgjengelig fra: https://shap.readthedocs.io/en/latest/example_notebooks/api_examples/plots/beeswarm.html (Hentet 10. april 2023).
- Lundberg, S., 2018b. waterfall plots Tilgjengelig fra: https://shap.readthedocs.io/en/latest/example_notebooks/api_examples/plots/waterfall.html (Hentet: 20. april 2023).
- Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.I., 2020. From local explanations to global understanding with explainable ai for trees. *Nature Machine Intelligence* 2, 56–67. doi:10.1038/s42256-019-0138-9.
- Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems* 30.
- Mansi, S.A., Maxwell, W.F., Zhang, A.J., 2012. Bankruptcy prediction models and the cost of debt. *The Journal of Fixed Income* 21, 25–42.
- Marcílio, W.E., Eler, D.M., 2020. From explanations to feature selection: assessing shap values as feature selection mechanism, i: 2020 33rd SIBGRAPI conference on Graphics, Patterns and Images (SIBGRAPI), IEEE. s. 340–347. doi:10.1109/SIBGRAPI51738.2020.00053.
- Miller, T., 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 267, 1–38. doi:10.1016/j.artint.2018.07.007.
- Min, J.H., Jeong, C., 2009. A binary classification method for bankruptcy prediction. *Expert Systems with Applications* 36, 5256–5263. doi:10.1016/j.eswa.2008.06.073.
- Molnar, C., Konig, G., Herbinger, J., Freiesleben, T., Dandl, S., Scholbeck, C.A., Casalicchio, G., Grosse-Wentrup, M., Bischl, B., 2022. General pitfalls of model-agnostic interpretation methods for machine learning models, i: xxAI-Beyond Explainable AI: International Workshop, Held in Conjunction with ICML 2020,

-
- July 18, 2020, Vienna, Austria, Revised and Extended Papers, Springer. s. 39–68. doi:10.1007/978-3-031-04083-2_4.
- Murdoch, W.J., Singh, C., Kumbier, K., Abbasi-Asl, R., Yu, B., 2019. Interpretable machine learning: definitions, methods, and applications. arXiv preprint arXiv:1901.04592 doi:10.48550/arXiv.1901.04592.
- Næringsdepartementet, 2019. Regjeringen besøker småbedrifter over hele landet i juni. Regjeringen.no Tilgjengelig fra: <https://www.regjeringen.no/no/dokumentarkiv/regjeringen-solberg/aktuelt-regjeringen-solberg/nfd/pressemeldinger/2019/ny-side17/id2661499/> (Hentet 20. mars 2023).
- Ohlson, J.A., 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research* , 109–131doi:10.2307/2490395.
- Olson, D.L., Delen, D., Meng, Y., 2012. Comparative analysis of data mining methods for bankruptcy prediction. *Decision Support Systems* 52, 464–473. doi:10.1016/j.dss.2011.10.007.
- Ouma, O.P., mfl., 2012. The relationship between dividend payout and firm performance: A study of listed companies in kenya. *European scientific journal* 8.
- Papik, M., Lenka, P., 2023. Impacts of crisis on sme bankruptcy prediction models' performance. *Expert Systems with Applications* 214, 1–17. doi:10.1016/j.eswa.2022.119072.
- Paraschiv, F., Schmid, M., Wahlstrøm, R.R., 2021. Bankruptcy prediction of privately held SMEs using feature selection methods. *SSRN Electronic Journal* doi:10.2139/ssrn.3911490.
- Park, M.S., Son, H., Hyun, C., Hwang, H.J., 2021. Explainability of machine learning models for bankruptcy prediction. *IEEE Access* 9, 124887–124899. doi:10.1109/ACCESS.2021.3110270.
- Pereira, J.M., Basto, M., Da Silva, A.F., 2016. The logistic lasso and ridge regression in predicting corporate failure. *Procedia Economics and Finance* 39, 634–641. doi:10.1016/S2212-5671(16)30310-0.
- Premachandra, I.M., Bhabra, G.S., Sueyoshi, T., 2009. Dea as a tool for bankruptcy assessment: A comparative study with logistic regression technique. *European*

-
- Journal of Operational Research 193, 412–424. doi:10.1016/j.ejor.2007.11.036.
- Qian, H., Wang, B., Yuan, M., Gao, S., Song, Y., 2022. Financial distress prediction using a corrected feature selection measure and gradient boosted decision tree. *Expert Systems with Applications* 190, 116202. doi:10.1016/j.eswa.2021.116202.
- Regjeringen, 2022. Støtteordninger for næringsliv og kultur Tilgjengelig fra: <https://www.regjeringen.no/no/tema/naringsliv/stotteordninger-for-naeringsliv-og-kultur/id2694423/?expand=factbox2841769> (Hentet: 14. april 2023).
- Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* 1, 206–215. doi:10.1038/s42256-019-0048-x.
- Russell, S.J., 2010. *Artificial Intelligence a Modern Approach*. Pearson Education, Inc., New Jersey.
- Rybalka, M., 2021. Stor økning i direkte støtte til næringslivet under korona Tilgjengelig fra: <https://www.ssb.no/teknologi-og-innovasjon/forskning-og-innovasjon-i-naeringslivet/statistikk/naeringspolitiske-virkemidler/artikler/stor-okning-i-direkte-stotte-til-naeringslivet-under-korona> (Hentet: 14. april 2023).
- Samuel, A.L., 1959. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development* 3, 210–229. doi:10.1147/rd.33.0210.
- Sanchez-Marono, N., Alonso-Betanzos, A., Tombilla-Sanromán, M., 2007. Filter methods for feature selection—a comparative study. *Lecture notes in computer science* 4881, 178–187.
- Saraswat, M., 2016. Beginners tutorial on xgboost and parameter tuning in r Tilgjengelig fra: <https://www.hackerearth.com/practice/machine-learning/machine-learning-algorithms/beginners-tutorial-on-xgboost-parameter-tuning-r/tutorial/> (Hentet: 12. april 2023).
- Schalck, C., Yankol-Schalck, M., 2021. Predicting french sme failures: new evidence from machine learning techniques. *Applied Economics* 53, 5948–5963. doi:10.1080/00036846.2021.1934389.

-
- Schapiro, R.E., 2013. Explaining adaboost. *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*, 37–52doi:10.1007/978-3-642-41136-6_5.
- Sen, P.C., Hajra, M., Ghosh, M., 2020. Supervised classification algorithms in machine learning: A survey and review, i: *Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph 2018*, Springer. s. 99–111.
- Shi, Y., Li, X., 2019. An overview of bankruptcy prediction models for corporate firms: A systematic literature review. *Intangible Capital* 15, 114–127. doi:10.3926/ic.1354.
- Shumway, T., 2001. Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business* 74, 101–124. doi:10.1086/209665.
- Singh, A., Thakur, N., Sharma, A., 2016. A review of supervised machine learning algorithms, i: *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, Ieee. s. 1310–1315.
- Sivam, R., Velten, C., Strøm, T., 2018. *Norske bedrifters verdigrunnlag*, 1–27.
- Slack, D., Hilgard, S., Jia, E., Singh, S., Lakkaraju, H., 2020. Fooling lime and shap: Adversarial attacks on post hoc explanation methods, i: *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, s. 180–186.
- Solorio-Fernandez, S., Carrasco-Ochoa, J.A., Martinez-Trinidad, J.F., 2016. A new hybrid filter–wrapper feature selection method for clustering based on ranking. *Neurocomputing* 214, 866–880. doi:10.1016/j.neucom.2016.07.026.
- Son, H., Hyun, C., Phan, D., Hwang, H.J., 2019. Data analytic approach for bankruptcy prediction. *Expert Systems with Applications* 138, 112816. doi:10.1016/j.eswa.2019.07.033.
- Statistisk-sentralbyrå, 2022a. Kredittindikator Tilgjengelig fra: <https://www.ssb.no/statbank/table/06718/tableViewLayout1/> (Hentet: 27. april 2023).
- Statistisk-sentralbyrå, 2022b. Opna konkursar Tilgjengelig fra: <https://www.ssb.no/statbank/table/09694/tableViewLayout1/> (Hentet: 27. april 2023).
- Statistisk-sentralbyrå, 2023. Virksomheter Tilgjengelig fra: <https://www.ssb.no/virksomheter-foretak-og-regnskap/virksomheter-og-foretak/statistikk/virksomheter> (Hentet: 24. april 2023).
-

-
- Stein, R.M., 2005. The relationship between default prediction and lending profits: Integrating roc analysis and loan pricing. *Journal of Banking & Finance* 29, 1213–1236. doi:10.1016/j.jbankfin.2004.04.008.
- Stock, J.H., Watson, M.W., 1992. A procedure for predicting recessions with leading indicators: Econometric issues and recent experience doi:10.3386/w4014.
- Succurro, M., Mannarino, L., mfl., 2014. The impact of financial structure on firms' probability of bankruptcy: a comparison across western europe convergence regions. *Regional and Sectoral Economic Studies* 14, 81–94.
- Taffler, R.J., 1983. The assessment of company solvency and performance using a statistical model. *Accounting and Business Research* 13, 295–308. doi:10.1080/00014788.1983.9729767.
- Taffler, R.J., 1984. Empirical models for the monitoring of uk corporations. *Journal of Banking & Finance* 8, 199–227. doi:10.1016/0378-4266(84)90004-9.
- Tian, S., Yu, Y., 2017. Financial ratios and bankruptcy predictions: An international evidence. *International Review of Economics & Finance* 51, 510–526. doi: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. doi:10.1016/j.jbankfin.2014.12.003.
- Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58, 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x.
- Tobback, E., Bellotti, T., Moeyersoms, J., Stankova, M., Martens, D., 2017. Bankruptcy prediction for smes using relational data. *Decision Support Systems* 102, 69–81. doi:10.1016/j.dss.2017.07.004.
- Tran, K.L., Le, H.A., Nguyen, T.H., Nguyen, D.T., 2022. Explainable machine learning for financial distress prediction: Evidence from vietnam. *Data* 7, 160. doi:10.3390/data7110160.
- Trinkle, B.S., Baldwin, A.A., 2007. Interpretable credit model development via artificial neural networks. *Intelligent Systems in Accounting, Finance & Management: International Journal* 15, 123–147. doi:10.1002/isaf.289.

-
- Tsai, C.F., Hsu, Y.F., Yen, D.C., 2014. A comparative study of classifier ensembles for bankruptcy prediction. *Applied Soft Computing* 24, 977–984. doi:10.1016/j.asoc.2014.08.047.
- Vallini, C., Ciampi, F., Gordini, N., Benvenuti, M., 2008. Can credit scoring models effectively predict small enterprise default? statistical evidence from italian firms, i: Proceedings of the 8th Global Conference on Business & Economics, Association for Business and Economics Research (ABER), s. 1–23.
- Van Gestel, T., Baesens, B., Suykens, J., Espinoza, M., Baestaens, D.E., Vant-hienen, J., De Moor, B., 2003. Bankruptcy prediction with least squares support vector machine classifiers, i: 2003 IEEE International Conference on Computational Intelligence for Financial Engineering, 2003. Proceedings., IEEE. s. 1–8. doi:10.1109/CIFER.2003.1196234.
- Veganzones, D., Séverin, E., 2018. An investigation of bankruptcy prediction in imbalanced datasets. *Decision Support Systems* 112, 111–124. doi:10.1016/j.dss.2018.06.011.
- Wahlstrøm, R.R., 2022. Financial statements of companies in Norway doi:10.48550/ARXIV.2203.12842.
- Wahlstrøm, R.R., 2023. Explainable artificial intelligence (XAI) for interpreting machine learning methods and their individual predictions. Available at SSRN 4321303 doi:10.2139/ssrn.4321303.
- Wanner, J., Heinrich, K., Janiesch, C., Zschech, P., 2020. How much ai do you require? decision factors for adopting ai technology., i: ICIS.
- Wanner, J., Herm, L.V., Heinrich, K., Janiesch, C., 2021. Stop ordering machine learning algorithms by their explainability! an empirical investigation of the tradeoff between performance and explainability, i: Responsible AI and Analytics for an Ethical and Inclusive Digitized Society: 20th IFIP WG 6.11 Conference on e-Business, e-Services and e-Society, I3E 2021, Galway, Ireland, September 1–3, 2021, Proceedings 20, Springer. s. 245–258. doi:10.1016/j.ijinfomgt.2022.102538.
- Xia, Y., Liu, C., Li, Y., Liu, N., 2017. A boosted decision tree approach using bayesian hyper-parameter optimization for credit scoring. *Expert systems with applications* 78, 225–241. doi:10.1016/j.eswa.2017.02.017.

-
- Xiaomao, X., Xudong, Z., Yuanfang, W., 2019. A comparison of feature selection methodology for solving classification problems in finance, i: *Journal of Physics: Conference Series*, IOP Publishing. s. 6. doi:10.1088/1742-6596/1284/1/012026.
- Yildirim, M., Okay, F.Y., Özdemir, S., 2021. Big data analytics for default prediction using graph theory. *Expert Systems with Applications* 176, 114840. doi:10.1016/j.eswa.2021.114840.
- Zednik, C., 2021. Solving the black box problem: A normative framework for explainable artificial intelligence. *Philosophy & Technology* 34, 265–288. doi:10.1007/s13347-019-00382-7.
- Zerilli, J., Knott, A., Maclaurin, J., Gavaghan, C., 2019. Transparency in algorithmic and human decision-making: Is there a double standard? *Philosophy & Technology* 32, 661–683. doi:10.1007/s13347-018-0330-6.
- Zhang, G., Hu, M.Y., Patuwo, B.E., Indro, D.C., 1999. Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research* 116, 16–32. doi:10.1016/S0377-2217(98)00051-4.
- Zhou, J., Gandomi, A.H., Chen, F., Holzinger, A., 2021. Evaluating the quality of machine learning explanations: A survey on methods and metrics. *Electronics* 10, 593. doi:10.3390/electronics10050593.
- Zhou, L., 2013. Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods. *Knowledge-Based Systems* 41, 16–25. doi:10.1016/j.knosys.2012.12.007.
- Zikeba, M., Tomczak, S.K., Tomczak, J.M., 2016. Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. *Expert Systems with Applications* 58, 93–101. doi:10.1016/j.eswa.2016.04.001.
- Zmijewski, M.E., 1984. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting research* , 59–82doi:10.2307/2490859.

Appendix

A Regnskapsposter i opprinnelig datasett

Tabell 8: Regnskapsposter i det opprinnelige datasettet

Regnskapsposter i opprinnelig datasett	
1. Orgnr	48. Kundefordringer
2. Regnaar	49. Andre kortsiktige fordringer
3. Startdato	50. Krav på innbetaling av selskapskapital
4. Avslutningsdato	51. Sum fordringer
5. Mottakstype	52. Investeringer
6. Avviklingsregnskap	53. Aksjer og andeler i samme konsern
7. Feilvalgør	54. Markedsbaserte aksjer
8. Orgform	55. Markedsbaserte obligasjoner
9. Mor_i_konsern	56. Andre markedsbaserte finansielle instrumenter
10. Regler_små	57. Andre kortsiktige finansielle instrumenter
11. Journalnr	58. Sum investeringer
12. Ifrs_selskap	59. Kasse/Bank/Post
13. Fravalg_revisjon	60. Sum omløpsmidler
14. Utarbeidet_regnskapsfører	61. Sum eiendeler
15. Bistand_regnskapsfører	62. Egenkapital og gjeld
16. Eiendeler	63. Egenkapital
17. Anleggsmidler	64. Innskutt egenkapital
18. Immaterielle eiendeler	65. Selskapskapital
19. Forskning og utvikling	66. Egne aksjer
20. Konesjoner, patenter, lisenser, varemerker og lignende rettigheter	67. Overkursfond
21. Utsatt skattefordel	68. Annen innskutt egenkapital
22. Goodwill	69. Sum innskutt egenkapital
23. Øvrige immaterielle eiendeler (jfr. Verdi leiekontrakter m.m.)	70. Opptjent egenkapital
24. Sum immaterielle anleggsmidler	71. Fond for vurderingsforskjeller
25. Varige driftsmidler	72. Annen egenkapital
26. Tomter, bygninger og annen fast eiendom	73. Udekket tap
27. Maskiner og anlegg	74. Sum opptjent egenkapital
28. Skip, rigger, fly og lignende	75. Sum egenkapital
29. Driftsløsøre, inventar, verktøy, kontormaskiner og lignende	76. Minoritetsinteresser* Konsernpost
30. Øvrige driftsmidler	77. Gjeld
31. Sum varige driftsmidler	78. Avsetninger for forpliktelser
32. Finansielle anleggsmidler	79. Pensjonsforpliktelser
33. Investeringer i datterselskap	80. Utsatt skatt
34. Investeringer i annet foretak i samme konsern	81. Andre avsetninger for forpliktelser
35. Lån til foretak i samme konsern	82. Sum avsetninger til forpliktelser
36. Investeringer i tilknyttet selskap	83. Annen langsiktig gjeld
37. Lån til tilknyttet selskap og felles kontrollert virksomhet	84. Konvertible lån
38. Investeringer i aksjer og andeler	85. Obligasjonslån
39. Obligasjoner og andre langsiktige fordringer	86. Gjeld til kredittinstitusjoner
40. Sum finansielle anleggsmidler	87. Øvrig langsiktig gjeld
41. Sum anleggsmidler	88. Sum langsiktig gjeld
42. Omløpsmidler	89. Kortsiktig gjeld
43. Varer	90. Sertifikatlån
44. Råvarer	91. Leverandørgjeld
45. Varer i arbeid	92. Betalbar skatt
46. Ferdigvarer	93. Skyldige offentlige utgifter
47. Fordringer	94. Utbytte
	95. Annen kortsiktig gjeld
	96. Sum kortsiktig gjeld
	97. Sum egenkapital og gjeld
	98. Salgsinntekter
	99. Honorarer

100.	Offentlig avgift vedr. omsetning (alkoholavg, pant mv)	140.	Tap på fordring
101.	Offentlig tilskudd/refusjon	141.	Frakt
102.	Uopptjent inntekt	142.	Energi
103.	Annen driftsinntekt	143.	Kostnad lokaler
104.	Leieinntekter	144.	Leie maskin, inventar o.l.
105.	Provisjons- og kommisjonsinntekt	145.	Reparasjon og vedlikehold
106.	Gevinst ved avgang anleggsmidler	146.	Kostnader reiser
107.	Annen driftsrelatert inntekt (inkl. abonnement / reklameinntekter etc.)	147.	Forsikringer
108.	Fraktinntekter	148.	Andre driftskostnader (fou kostnader, adm.kostnader, markedsfoeringskostnader)
109.	Inntekt på investering i datterselskap	149.	Sum driftskostnader
110.	Gevinst ved salg av verdipapirer	150.	Driftsresultat
111.	Sum driftsinntekter	151.	Konsernbidrag
112.	Endring i beholdning av varer under tilvirkning og ferdig tilvirkede varer	152.	Resultatandel datterselskap
113.	Endring i beholdning av varer under tilvirkning	153.	Inntekt på investering i annet foretak i samme konsern
114.	Endring i beholdning ferdig tilvirkende varer	154.	Resultatandel annet foretak i samme konsern
115.	Beholdningsendring av egentilvirkede anleggsmidler	155.	Inntekt på investering i tilknyttet selskap
116.	Varekostnad	156.	Resultatandel tilknyttet selskap
117.	Forbruk av råvarer og innkjøpte halvfabrikata	157.	Renteinntekter fra foretak i samme konsern
118.	Forbruk av varer under tilvirkning	158.	Annen renteinntekt
119.	Forbruk av ferdig tilvirkede varer	159.	Renteinntekt skattefri
120.	Forbruk av innkjøpte varer for videresalg	160.	Annen finansinntekt
121.	Fremmedytelse og underentreprise	161.	Realisert valutagevinster (Agio)
122.	Annen periodisering	162.	Urealisert valutagevinster (Agio)
123.	Lønnskostnader	163.	Sum finansinntekter
124.	Lønn til ansatte	164.	Verdiendring markedsbaserte finansielle omløpsmidler
125.	Fordeler i arbeidsforhold	165.	Verdiøkning markedsbaserte finansielle omløpsmidler
126.	Annen oppgavepliktig godtgjørelse	166.	Verdireduksjon markedsbaserte finansielle omløpsmidler
127.	Arbeidsgiveravgift og pensjonskostnad	167.	Nedskrivning av andre finansielle omløpsmidler
128.	Annen kostnadsgodtgjørelse	168.	Nedskrivning av finansielle anleggsmidler
129.	Arbeidsgodtgjørelse til eiere o.l.	169.	Rentekostnader til foretak i samme konsern
130.	Offentlig tilskudd vedr. arbeidskraft	170.	Annen rentekostnad
131.	Offentlig refusjon vedr. arbeidskraft	171.	Rentekostnad, ikke fradragsberettiget
132.	Annen personalkostnad	172.	Annen finanskostnad
133.	Avskrivning på varige driftsmidler og immaterielle eiendeler	173.	Realisert valutatap (disagio)
134.	Avskrivninger på bygninger og annen fast eiendom	174.	Urealisert valutatap (disagio)
135.	Avskrivninger på transportmidler, maskiner og inventar	175.	Sum finanskostnader
136.	Avskrivning på immaterielle eiendeler	176.	Netto finans
137.	Avskrivning på varige driftsmidler og immaterielle eiendeler	177.	Ordinært resultat før skattekostnad
138.	Nedskrivning av varige driftsmidler og immaterielle eiendeler	178.	Skattekostnad på ordinært resultat
139.	Annen driftskostnad	179.	Refusjon skatt etter aksjeloven
		180.	Endring utsatt skatt

181.	Ordinært resultat	227.	Rentekostnad til foretak i samme konsern
182.	Ekstraordinære inntekter	228.	Ordinært resultat etter skattekostnad
183.	Ekstraordinære kostnader	229.	Ekstraordinær inntekt
184.	Netto ekstraordinære poster	230.	Ekstraordinær kostnad
185.	Skattekostnad på ekstraordinært resultat	231.	Overføring til/fra fond for vurderingsforskjeller
186.	Årsresultat	232.	Fondsemisjon
187.	Overføringer og disponeringer	233.	Andre immaterielle eiendeler
188.	Overføringer til/fra fond for vurderingsforskjeller	234.	Fond for verdiendringer
189.	Avsatt utbytte	235.	Sum gjeld
190.	Aksjonærbidrag	236.	Resultat av ekstraordinære poster
191.	Overføringer til/fra annen egenkapital	237.	Årsresultat før minoritetsinteresser
192.	Sum overføringer og disponeringer	238.	Investeringseiendom
193.	Sum immaterielle eiendeler	239.	Biologiske eiendeler
194.	Investering i datterselskap	240.	Sum varer
195.	Investering i annet foretak i samme konsern	241.	Fond for urealiserte gevinster
196.	Obligasjoner og andre fordringer	242.	Sum annen langsiktig gjeld
197.	Andre fordringer	243.	Sum egenkapital og gjeld
198.	Konsernfordringer	244.	Posten utenom balansen
199.	Aksjer og andeler i foretak i samme konsern	245.	Verdiøkning av markedsbaserte finansielle omløpsmidler
200.	Andre finansielle instrumenter	246.	Verdiøkning andre finansielle instrumenter vurdert til virkelig verdi
201.	Bankinnskudd, kontanter og lignende	247.	Verdireduksjon av markedsbaserte finansielle omløpsmidler
202.	Sum eiendeler	248.	Verdireduksjon andre finansielle instrumenter vurdert til virkelig verdi
203.	Beholdning av egne aksjer	249.	Overføring til/fra fond for urealiserte gevinster
204.	Annen Egenkapital	250.	Obligasjoner
205.	Minoritetsinteresser	251.	Fond
206.	Langsiktig gjeld	252.	Overføring til/fra fond
207.	Sum avsetninger for forpliktelser	253.	Sum bankinnskudd, kontanter og lignende
208.	Langsiktig konserngjeld	254.	Avsetning for forpliktelser
209.	Ansvarlig lånekapital	255.	Inntekt på investering i datterselskap og tilknyttet selskap
210.	Skyldige offentlige avgifter	256.	Inntekt på andre investeringer
211.	Kortsiktig konserngjeld	257.	Nedskrivning av finansielle eiendeler
212.	Sum egenkapital og gjeld	258.	Ekstraordinære poster
213.	Poster utenom balansen	259.	Konserfordringer
214.	Garantistillelser	260.	Finansinntekter og finanskostnader
215.	Pantstillelser	261.	Årsresultat etter minoritetsinteresser
216.	Inntekter	262.	BALANSE – EIENDELER
217.	Salgsinntekt	263.	BALANSE - EGENKAPITAL OG GJELD
218.	Sum inntekter	264.	Overkurs
219.	Kostnader	265.	Gjeld
220.	Endring i beholdning av egentilvirkede anleggsmidler	266.	RESULTATREGNSKAP
221.	Lønnskostnad	267.	Andre resultatkomponenter for IFRS-foretak
222.	Sum kostnader	268.	Totalresultat
223.	Finansinntekter	269.	Avgitt konsernbidrag
224.	Renteinntekt fra foretak i samme konsern		
225.	Finanskostnader		
226.	Verdiendring av markedsbaserte finansielle omløpsmidler		

270.	Sum resultatkomponenter for IFRS-foretak
271.	Ordinært utbytte
272.	Tilleggsutbytte
273.	Ekstraordinært utbytte
274.	Ikke registrert kapitalforhøyelse
275.	Skattekostnad på ekstraordinære poster
276.	Stiftelsesdato
277.	Næringskode
278.	Konkursdato
279.	Landkode
280.	Postnummer
281.	Sum_eiendeler_EUR
282.	Sum_omsetning_EUR
283.	Age_in_days
284.	Næringskoder_level_1
285.	Næringskoder_level_2
286.	Can_opt_out
287.	Bankrupt_fs

B Variabler i endelig datasett

Tabell 9: Liste over variabler som er benyttet i modellene

Variabler i endelig datasett	
1. (inventar + kundefordringer)/total egenkapital	47. total egenkapital/anleggsmidler
2. (langsiktig gjeld + total egenkapital)/anleggsmidler	48. total egenkapital/salg
3. kundefordringer/salg	49. intervall uten kreditt
4. omløpsmidler/kortsiktig gjeld	50. dummy; 1 hvis total gjeld er større enn totale eiendeler
5. (omløpsmidler/kortsiktig gjeld) * (driftsresultat/annen_rentekostnad)	51. driftskostnader/salg
6. netto inntjening/total egenkapital	52. kortsiktig likviditet/total gjeld
7. EBITDA/total gjeld	53. driftsresultat/total inntekter
8. total egenkapital/total gjeld	54. driftsresultat/innskutt egenkapital
9. kortsiktig likviditet som prosent av sysselsatt kapital	55. driftsmidler/total eiendeler
10. kortsiktig likviditet/salg	56. personalkostnader/merverdi
11. kortsiktig likviditet /kortsiktig gjeld	57. resultat før skatt / innskutt kapital
12. kortsiktig likviditet/total eiendeler	58. årsresultat /total inntekter
13. salg/omløpsmidler	59. overskudd/netto arbeidskapital
14. omløpsmidler/total egenkapital	60. omløpsmidler/salg
15. omløpsmidler/salg	61. omløpsmidler/total eiendeler
16. (omløpsmidler/total eiendeler) * (netto likvide eiendeler/total eiendeler)	62. resultat etter skatt og rentekostnader / netto sysselsatt kapital
17. kortsiktig gjeld/omløpsmidler	63. kortsiktig gjeld/resultat før skatt og rentekostnader
18. kortsiktig gjeld/total egenkapital	64. opptjent egenkapital/salg
19. kortsiktig gjeld/total gjeld	65. opptjent egenkapital/total eiendeler
20. kortsiktig gjeld/salg	66. avkastning på gjeld (inntjening/total gjeld)
21. total gjeld/total eiendeler	67. årsresultat/total eiendeler
22. kundefordringer/leverandørgjeld	68. totale inntekter/anleggsmidler
23. driftsresultat/(driftsresultat - rentekostnad)	69. totale inntekter/total eiendeler
24. EBIT/total eiendeler	70. totale inntekter/netto arbeidskapital
25. EBITDA/rentekostnad	71. salg/total eiendeler
26. effektiv skattesats	72. totale eiendeler/total inntekter
27. total egenkapital/total eiendeler	73. totale utgifter/total eiendeler
28. total egenkapital/langsiktig gjeld	74. totale inntekter/total utgifter
29. salg/total egenkapital	75. arbeidskapital/kortsiktig gjeld
30. resultat før skatt i prosent av sysselsatt kapital	76. arbeidskapital/salg
31. finansielle utgifter/salg	77. arbeidskapital/total eiendeler
32. EBIT/salg	78. arbeidskapital/total egenkapital
33. salg/anleggsmidler	79. dummy; 1 dersom innskutt kapital er mindre enn total egenkapital
34. anleggsmidler/total eiendeler	80. arbeidskapital/total inntekter
35. anleggsmidler/total egenkapital	81. leverandørgjeld/total eiendeler
36. immaterielle eiendeler/total eiendeler	82. skyldig offentlige avgifter/total eiendeler
37. rentekostnader/total inntekter	83. EBIT/total gjeld
38. rentebærende gjeld/total egenkapital	84. (rentefrie utgifter - lønn)/total eiendeler
39. inventar/kortsiktig gjeld	85. (aksjekapital + totale inntekter)/total eiendeler
40. inventar/arbeidskapital	86. salg/arbeidskapital
41. investeringsomsetning (salg/ (total egenkapital + total gjeld))	87. kortsiktig likviditet/omløpsmidler
42. total gjeld/total egenkapital	88. varekostnad/inventar
43. langsiktig gjeld/omløpsmidler	89. varekostnad/salg
44. årsresultat/innskutt egenkapital	90. (omløpsmidler - kortsiktig likviditet)/total eiendeler
45. årsresultat/salg	91. omløpsmidler/felles aksjonærkapital
46. (total inntekter-salg)/total inntekter	92. kortsiktig gjeld/total eiendeler
	93. utbytte/årsresultat

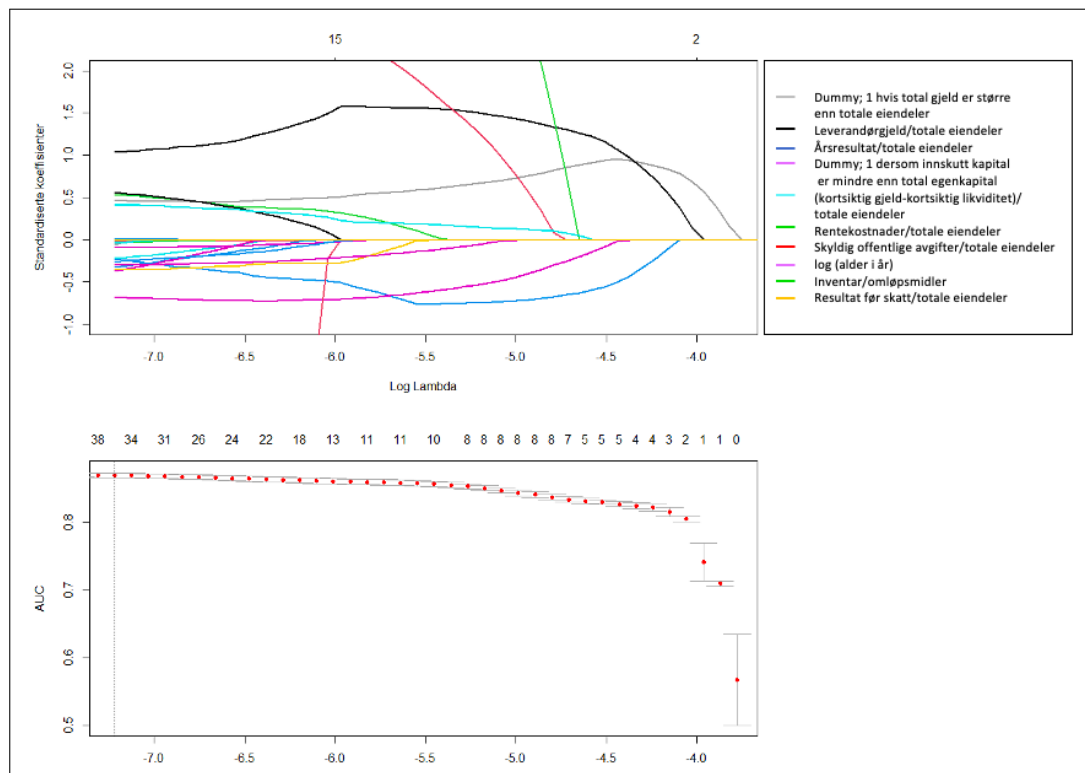
94. arbeidskapital/langsiktig gjeld	141. driftsinntekter/totale eiendeler
95. arbeidskapital/driftsutgifter	142. salgsinntekter/(sum_eiendeler - sum_kortsiktig_gjeld)
96. EBIT/totale immaterielle eiendeler	143. EBITDA/totale eiendeler
97. finansielle utgifter/salg	144. driftsresultat/totale eiendeler
98. anleggsmidler/(innskutt kapital + langsiktig gjeld)	145. driftsresultat/salg
99. (salg - varekostnad)/salg	146. (kortsiktig gjeld - kortsiktig likviditet)/totale eiendeler
100. inntektsgiring	147. leverandørgjeld/salg
101. immaterielle eiendeler/salg	148. opptjent egenkapital/kortsiktig gjeld
102. renteutgifter/total gjeld	149. (total egenkapital - immaterielle eiendeler) / (totale eiendeler - immaterielle eiendeler)
103. renteutgifter/totale utgifter	150. EBIT/rentekostnad
104. renteinntekter/renteutgifter	151. kundefordringer/total gjeld
105. renteinntekter/totale eiendeler	152. resultat før skatt/kortsiktig gjeld
106. inventar/varekostnad	153. omløpsmidler/total gjeld
107. inventar/omløpsmidler	154. log (alder i år)
108. inventar/salg	155. log (totale eiendeler)
109. langsiktig gjeld/total egenkapital	
110. langsiktig gjeld/totale eiendeler	
111. salg/immaterielle eiendeler	
112. årsresultat/bruttofortjeneste	
113. årsresultat / (total gjeld + innskutt kapital)	
114. netto omløpsmidler/inventar	
115. totale eiendeler / (total egenkapital + langsiktig gjeld)	
116. rentefri utgifter/driftsresultat	
117. totale inntekter/salg	
118. resultat før skatt/total egenkapital	
119. resultat før skatt/ordinære utgifter	
120. resultat før skatt/salg	
121. resultat før skatt/totale eiendeler	
122. aksjekapital/totale eiendeler	
123. leverandørgjeld/kortsiktig gjeld	
124. leverandørgjeld/varelager	
125. opptjent egenkapital/inventar	
126. opptjent egenkapital/immaterielle eiendeler	
127. avkastning på sysselsatt kapital	
128. avkastning på netto anleggsmidler	
129. lønn/total kapital	
130. salg/kortsiktig likviditet	
131. salg/varelager	
132. salg/kundefordringer	
133. salg/totale materielle eiendeler	
134. rentebærende gjeld/total gjeld	
135. andel av lønnskostnader	
136. (kortsiktige eiendeler - total gjeld)/total kapital	
137. soliditetsgrad	
138. salg/total egenkapital	
139. (totale inntekter + renteinntekt)/totale kostnader	
140. rentekostnader/totale eiendeler	

C Forward selection

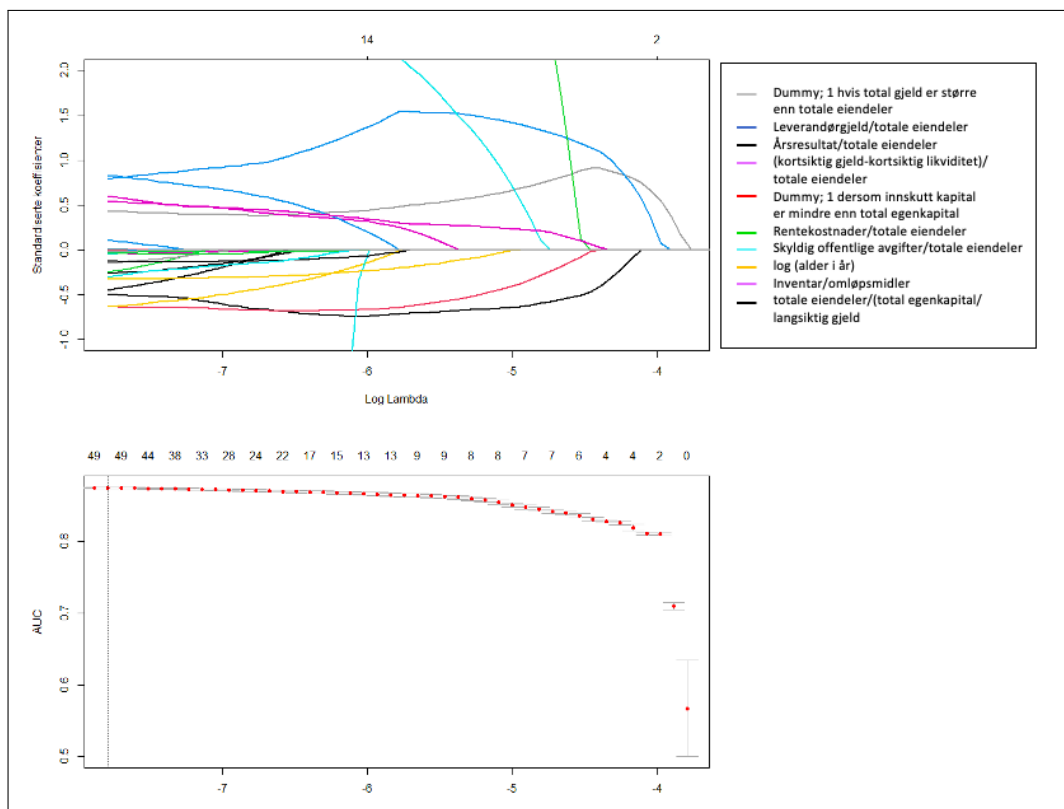
Tabell 10: Det benyttes rolling window for å dele dataen inn i trenings- og testsett. Treningssettene består av data fra fire år (de mørke blå rutene), mens testsettene består av data fra ett år (de lyse blå rutene). Foldene forskyves fremover i tid med ett år for hver ny fold. Videre så fremkommer antall observasjoner som har gått konkurs, ikke konkurs, og konkursfrekvensen.

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total
Fold															
1	Trening	Trening	Trening	Trening	Test										
2	Trening	Trening	Trening	Trening	Trening	Test									
3			Trening	Trening	Trening	Trening	Test								
4				Trening	Trening	Trening	Trening	Test							
5					Trening	Trening	Trening	Trening	Test						
6						Trening	Trening	Trening	Trening	Test					
7							Trening	Trening	Trening	Trening	Test				
8								Trening	Trening	Trening	Trening	Test			
9									Trening	Trening	Trening	Trening	Test		
10										Trening	Trening	Trening	Trening	Test	
<i>Obs.</i>	1 402	2 205	1 941	1 783	1 556	1 619	1 688	1 653	1 600	1 507	1 725	1 754	2 018	1 209	23 660
<i>Obs (ikke konkurs)</i>	96 134	101 781	104 871	105 589	107 290	109 988	114 523	118 968	123 344	127 816	132 133	137 314	141 784	146 711	1 668 246
<i>Konkurs-frekvens</i>	1,44%	2,12%	1,82%	1,66%	1,43%	1,45%	1,45%	1,37%	1,28%	1,17%	1,29%	1,26%	1,40%	0,82%	

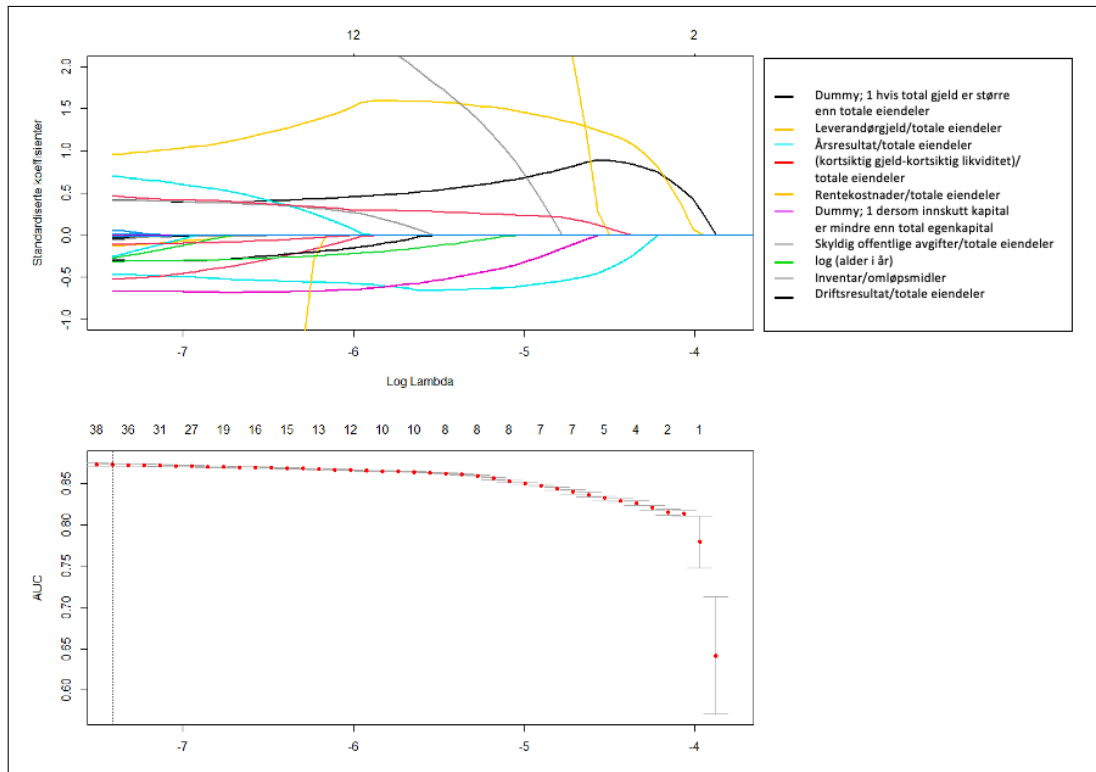
D LASSO stiplott



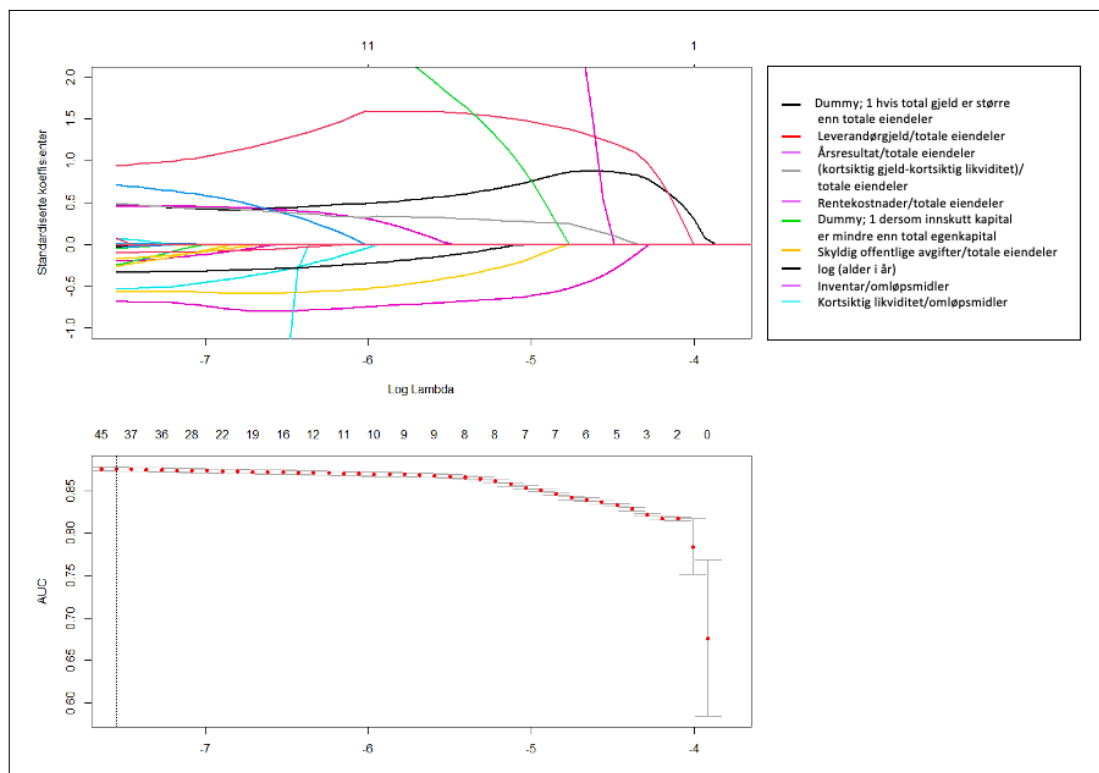
Figur 7: LASSO stiplott med alle firmaår for henholdsvis 2010 som testsett. Plottet viser verdier av λ fra en verdi som er så høy at alle estimerte koeffisienter settes til 0 (altså at ingen variabler blir valgt) til den verdien av λ som blir endelig valgt (se beskrivelse i kapittel 4.4.1). Under stiplottet er et tilhørende plott som viser AUC ved forskjellige verdier av λ . Med reduserende verdi av λ viser plottet at flere variabler blir valgt og at AUC øker. Aksen over plottet av AUC viser antall valgte variabler ved hver verdi av λ .



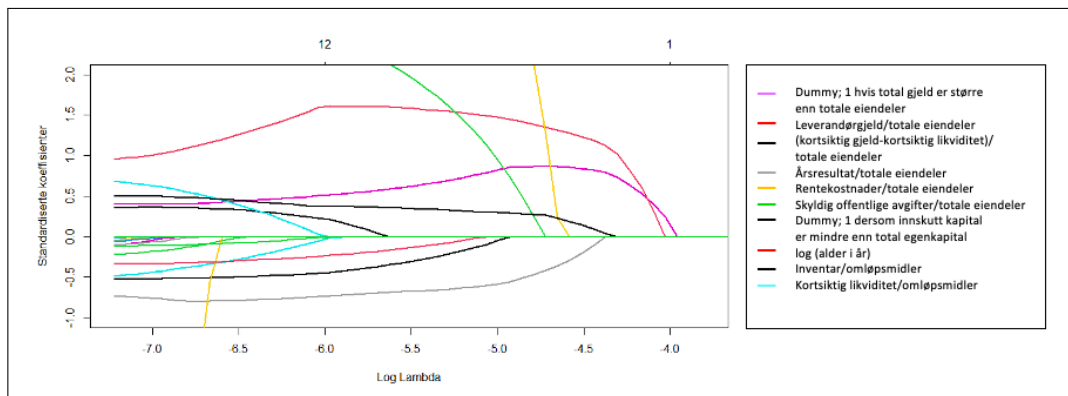
Figur 8: LASSO stiplott med alle firmaår for henholdsvis 2011 som testsett. Plottet viser verdier av λ fra en verdi som er så høy at alle estimerte koeffisienter settes til 0 (altså at ingen variabler blir valgt) til den verdien av λ som blir endelig valgt (se beskrivelse i kapittel 4.4.1). Under stiplottet er et tilhørende plott som viser AUC ved forskjellige verdier av λ . Med reduserende verdi av λ viser plottet at flere variabler blir valgt og at AUC øker. Aksen over plottet av AUC viser antall valgte variabler ved hver verdi av λ .



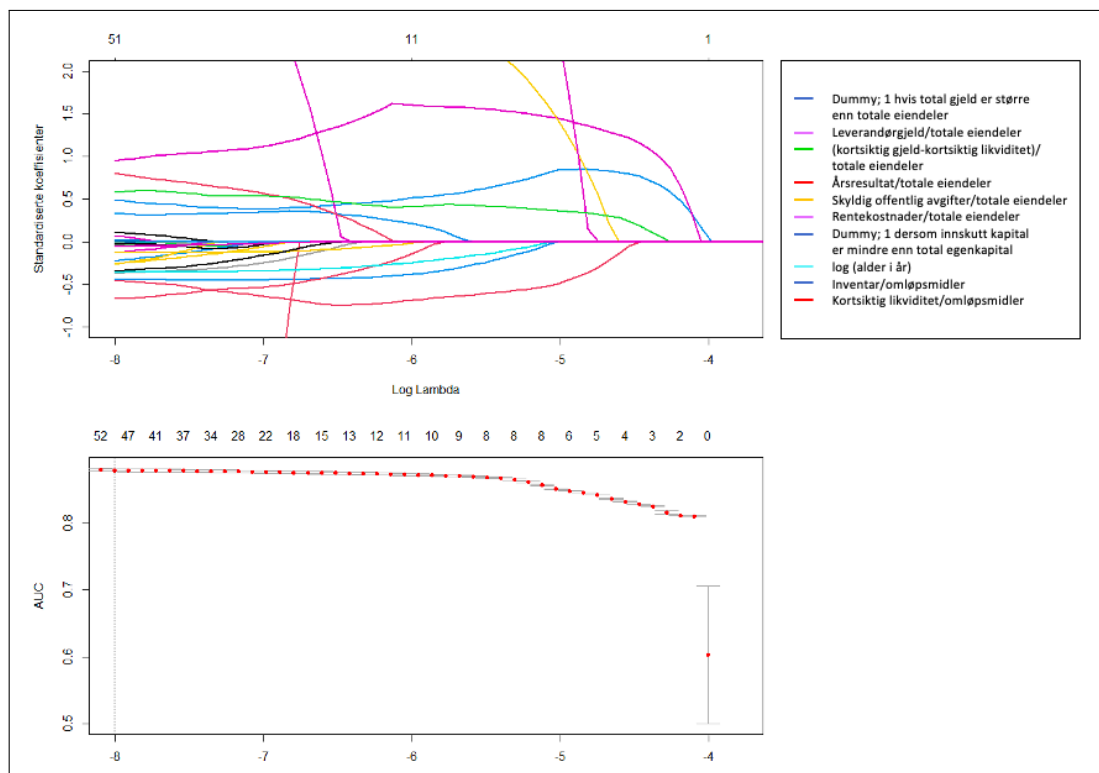
Figur 9: LASSO stiplott med alle firmaår for henholdsvis 2012 som testsett. Plottet viser verdier av λ fra en verdi som er så høy at alle estimerte koeffisienter settes til 0 (altså at ingen variabler blir valgt) til den verdien av λ som blir endelig valgt (se beskrivelse i kapittel 4.4.1). Under stiplottet er et tilhørende plott som viser AUC ved forskjellige verdier av λ . Med reduserende verdi av λ viser plottet at flere variabler blir valgt og at AUC øker. Aksen over plottet av AUC viser antall valgte variabler ved hver verdi av λ .



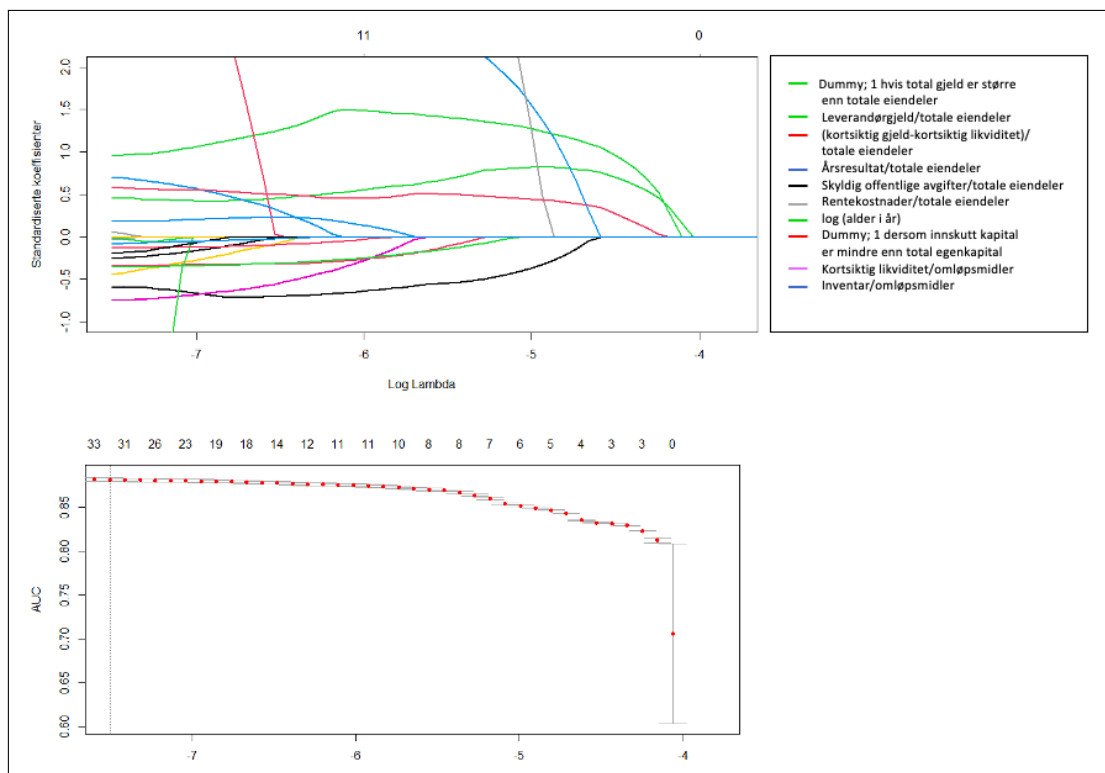
Figur 10: LASSO stiplott med alle firmaår for henholdsvis 2013 som testsett. Plottet viser verdier av λ fra en verdi som er så høy at alle estimerte koeffisienter settes til 0 (altså at ingen variabler blir valgt) til den verdien av λ som blir endelig valgt (se beskrivelse i kapittel 4.4.1). Under stiplottet er et tilhørende plott som viser AUC ved forskjellige verdier av λ . Med reduserende verdi av λ viser plottet at flere variabler blir valgt og at AUC øker. Aksen over plottet av AUC viser antall valgte variabler ved hver verdi av λ .



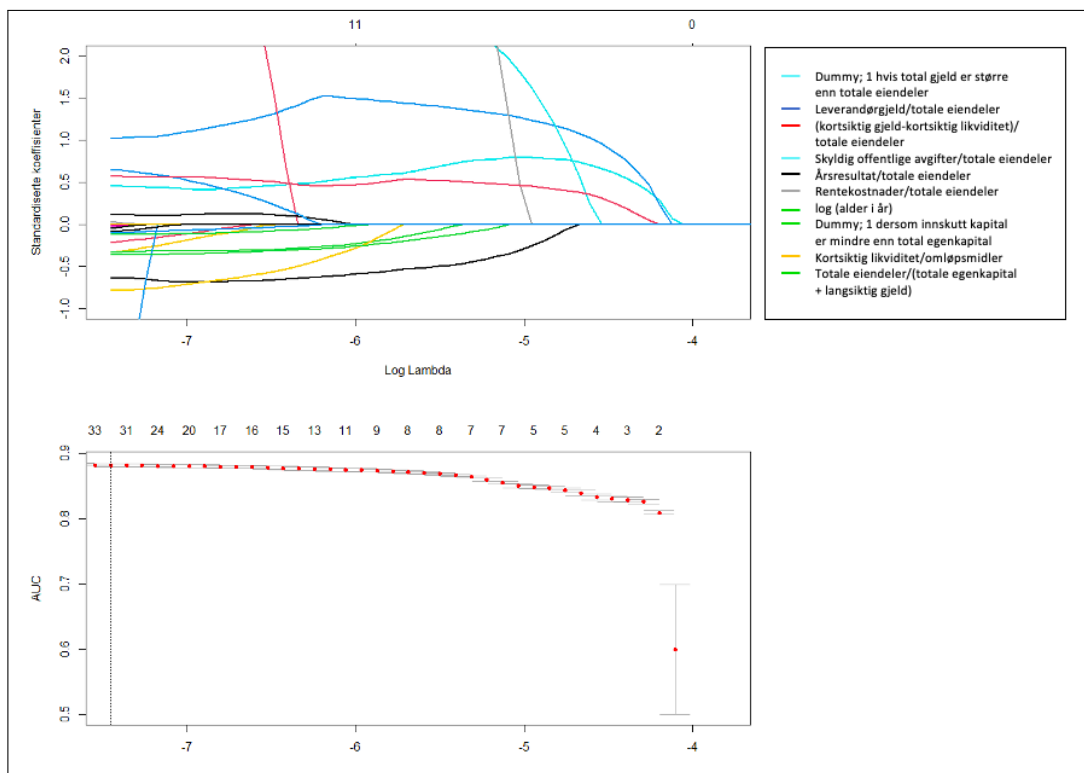
Figur 11: Stiplott for testår 2014. Her vises et LASSO-stiplott lagd ved bruk av den valgte lambdaverdi 0.0008. Til høyre vises en variabeloversikt over de ti viktigste variablene som LASSO har plukket ut. AUC-plott vises ikke her da vi har valgt en egen lambdaverdi 0.0008



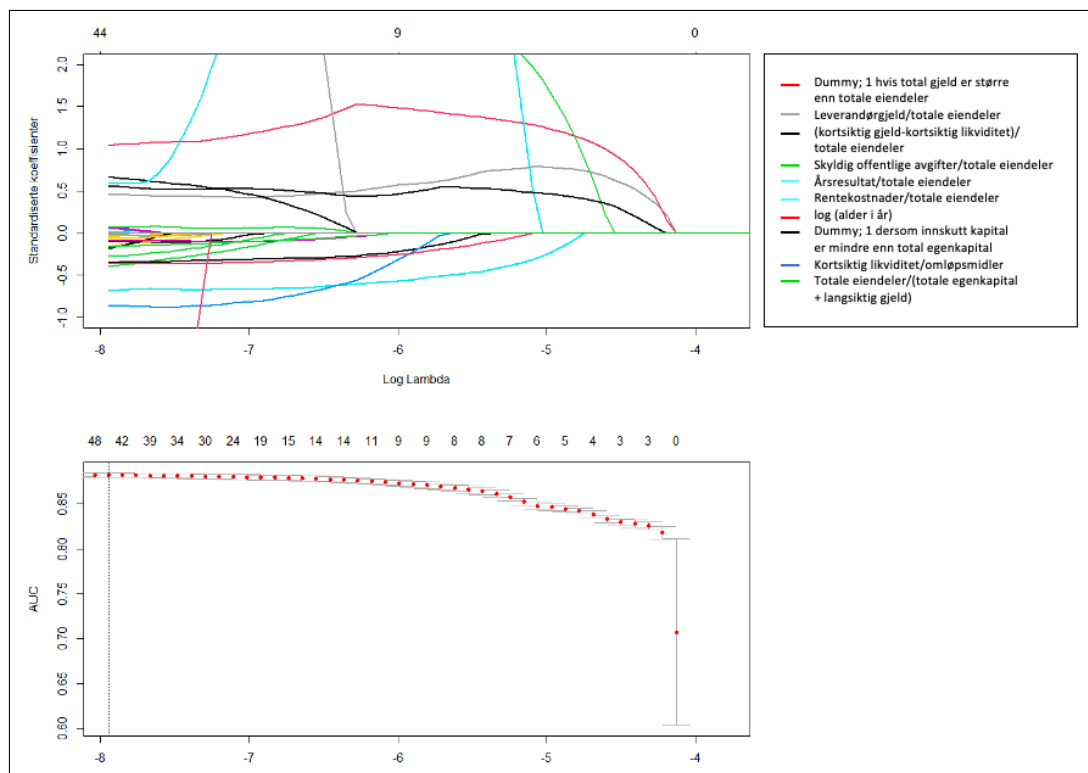
Figur 12: LASSO stiplott med alle firmaår for henholdsvis 2015 som testsett. Plottet viser verdier av λ fra en verdi som er så høy at alle estimerte koeffisienter settes til 0 (altså at ingen variabler blir valgt) til den verdien av λ som blir endelig valgt (se beskrivelse i kapittel 4.4.1). Under stiplottet er et tilhørende plott som viser AUC ved forskjellige verdier av λ . Med reduserende verdi av λ viser plottet at flere variabler blir valgt og at AUC øker. Aksen over plottet av AUC viser antall valgte variabler ved hver verdi av λ .



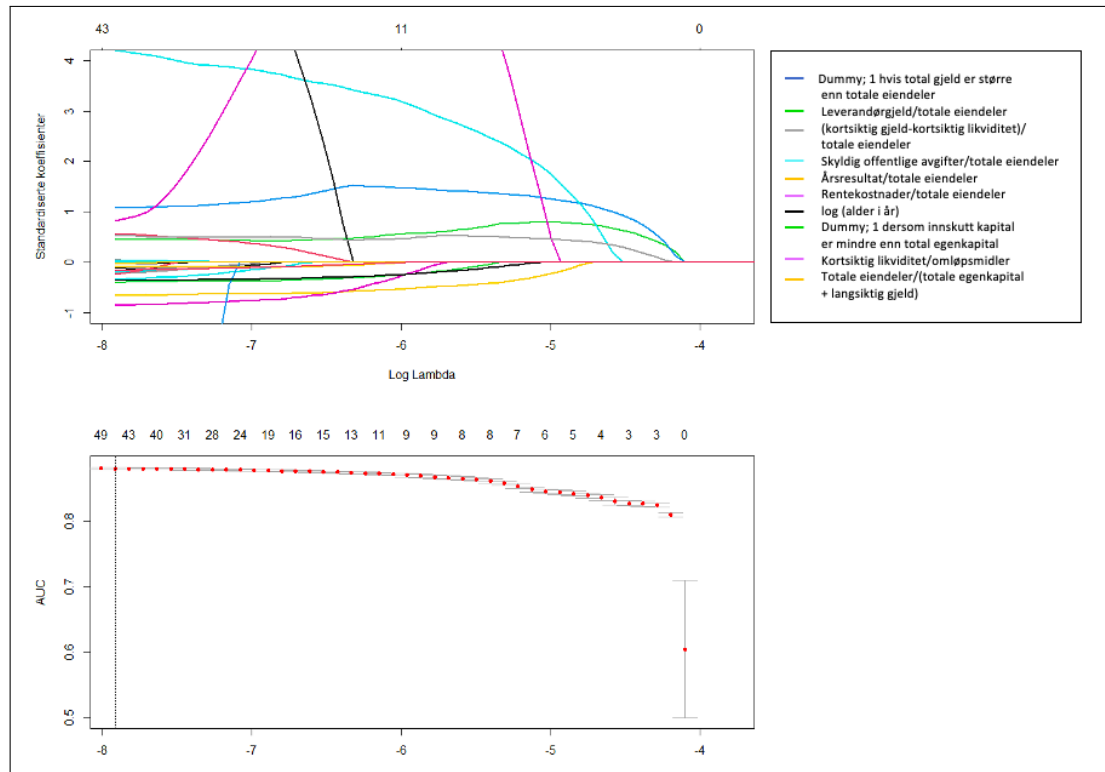
Figur 13: LASSO stiplott med alle firmaår for henholdsvis 2016 som testsett. Plottet viser verdier av λ fra en verdi som er så høy at alle estimerte koeffisienter settes til 0 (altså at ingen variabler blir valgt) til den verdien av λ som blir endelig valgt (se beskrivelse i kapittel 4.4.1). Under stiplottet er et tilhørende plott som viser AUC ved forskjellige verdier av λ . Med reduserende verdi av λ viser plottet at flere variabler blir valgt og at AUC øker. Aksen over plottet av AUC viser antall valgte variabler ved hver verdi av λ .



Figur 14: LASSO stiplott med alle firmaår for henholdsvis 2017 som testsett. Plottet viser verdier av λ fra en verdi som er så høy at alle estimerte koeffisienter settes til 0 (altså at ingen variabler blir valgt) til den verdien av λ som blir endelig valgt (se beskrivelse i kapittel 4.4.1). Under stiplottet er et tilhørende plott som viser AUC ved forskjellige verdier av λ . Med reduserende verdi av λ viser plottet at flere variabler blir valgt og at AUC øker. Aksen over plottet av AUC viser antall valgte variabler ved hver verdi av λ .

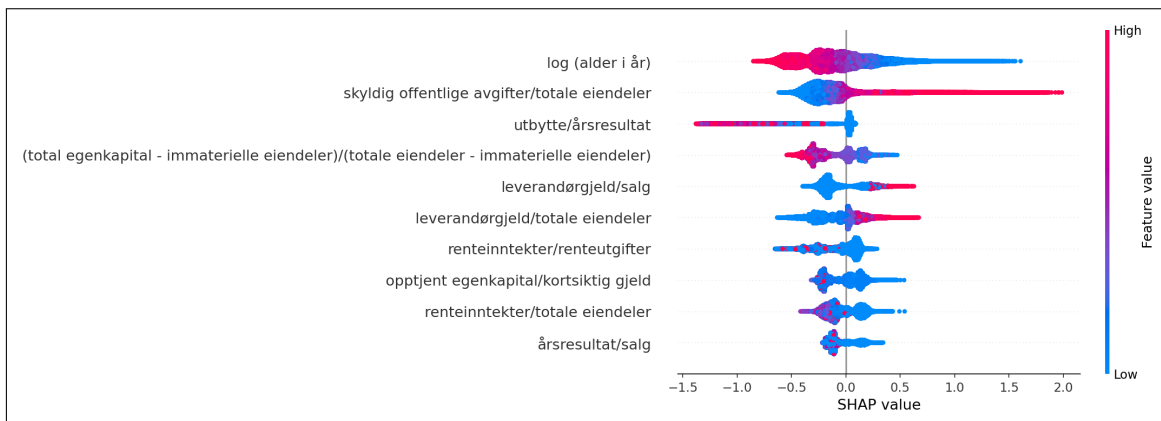


Figur 15: LASSO stiplott med alle firmaår for henholdsvis 2018 som testsett. Plottet viser verdier av λ fra en verdi som er så høy at alle estimerte koeffisienter settes til 0 (altså at ingen variabler blir valgt) til den verdien av λ som blir endelig valgt (se beskrivelse i kapittel 4.4.1). Under stiplottet er et tilhørende plott som viser AUC ved forskjellige verdier av λ . Med reduserende verdi av λ viser plottet at flere variabler blir valgt og at AUC øker. Aksene over plottet av AUC viser antall valgte variabler ved hver verdi av λ .

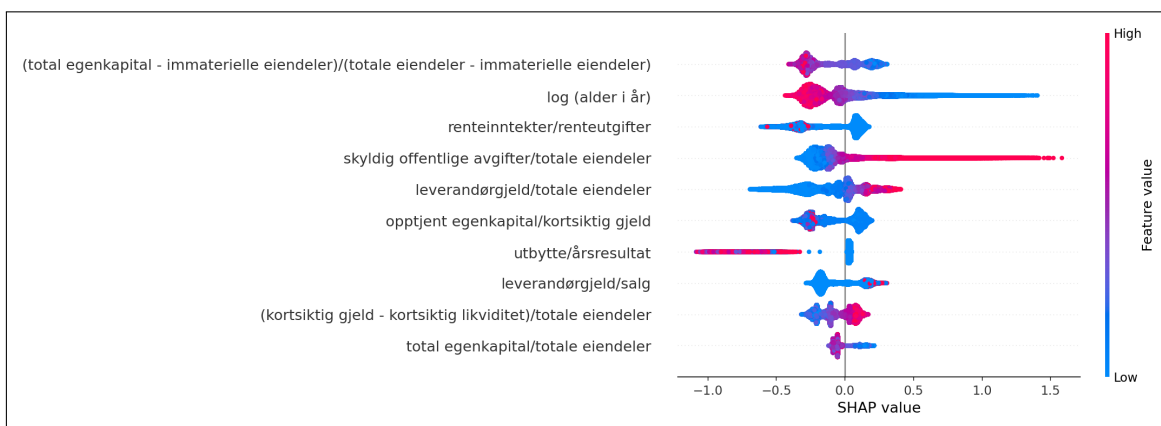


Figur 16: LASSO stiplott med alle firmaår for henholdsvis 2019 som testsett. Plottet viser verdier av λ fra en verdi som er så høy at alle estimerte koeffisienter settes til 0 (altså at ingen variabler blir valgt) til den verdien av λ som blir endelig valgt (se beskrivelse i kapittel 4.4.1). Under stiplottet er et tilhørende plott som viser AUC ved forskjellige verdier av λ . Med reduserende verdi av λ viser plottet at flere variabler blir valgt og at AUC øker. Aksen over plottet av AUC viser antall valgte variabler ved hver verdi av λ .

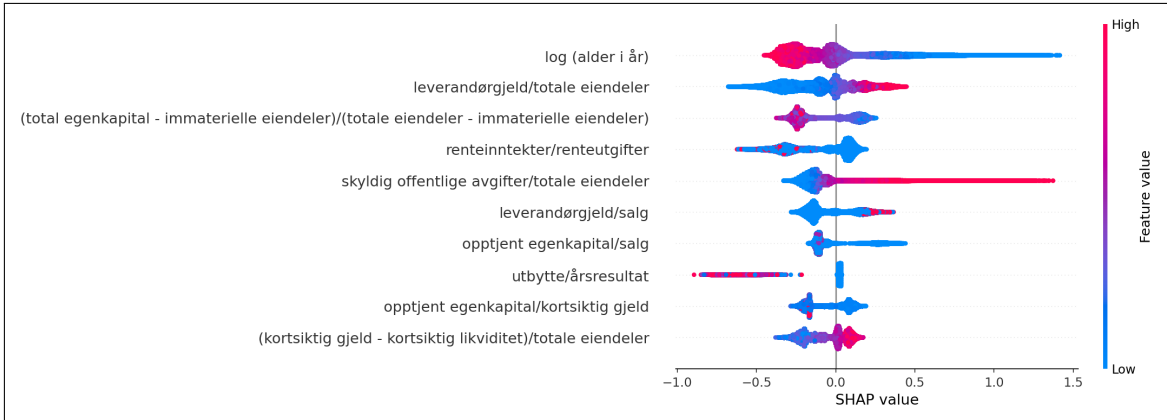
E SHAP beeswarm for hvert testår



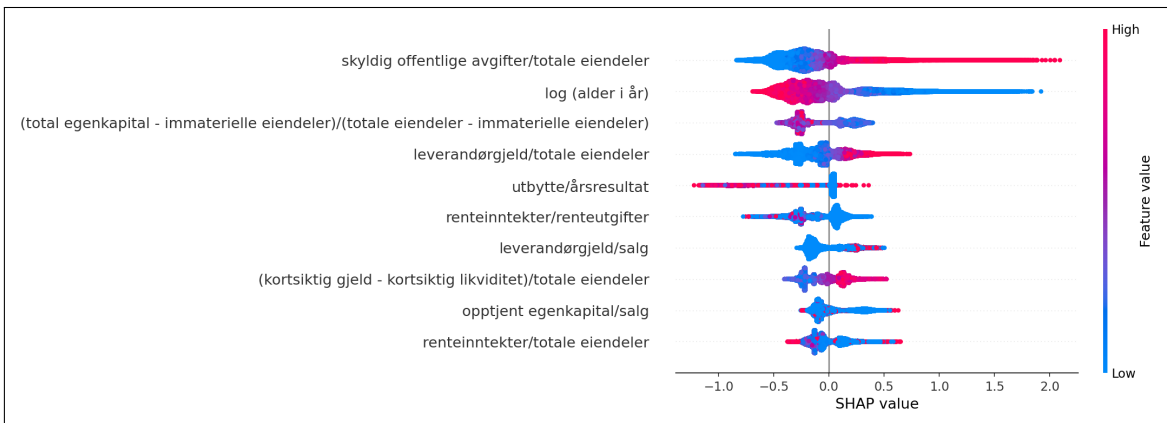
Figur 17: SHAP beeswarm plott for testår 2010. SHAP beesworm plottet viser den globale variabelviktigheten for de ti viktigste variablene i modellen for det gitte teståret. Den viktigste variabelen er øverst, deretter følge de andre variablene i synkende rekkefølge. Hver prikk i plottet representerer én observasjon. Fargen på prikkene viser hvorvidt verdien for variabelen er lav (blå) eller høy (rød). X-aksen representerer SHAP-verdien til observasjonen.



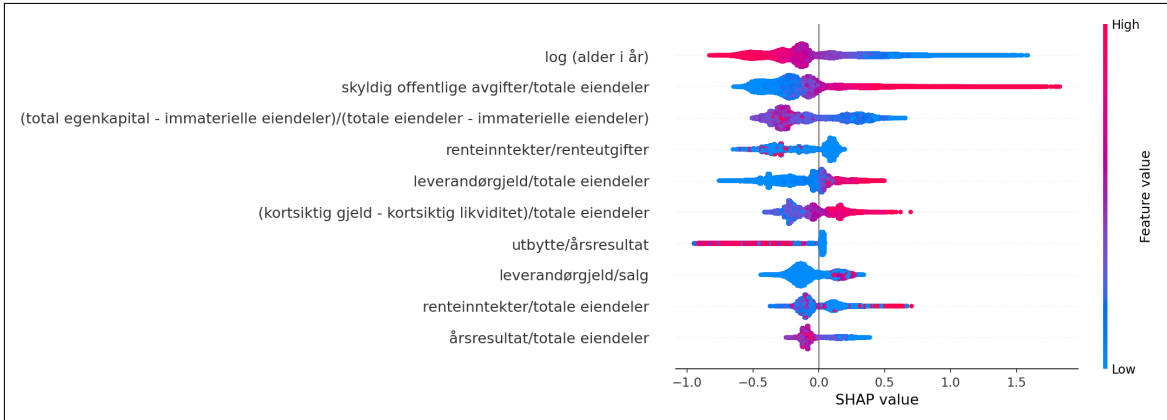
Figur 18: SHAP beeswarm plott for testår 2011. SHAP beesworm plottet viser den globale variabelviktigheten for de ti viktigste variablene i modellen for det gitte teståret. Den viktigste variabelen er øverst, deretter følge de andre variablene i synkende rekkefølge. Hver prikk i plottet representerer én observasjon. Fargen på prikkene viser hvorvidt verdien for variabelen er lav (blå) eller høy (rød). X-aksen representerer SHAP-verdien til observasjonen.



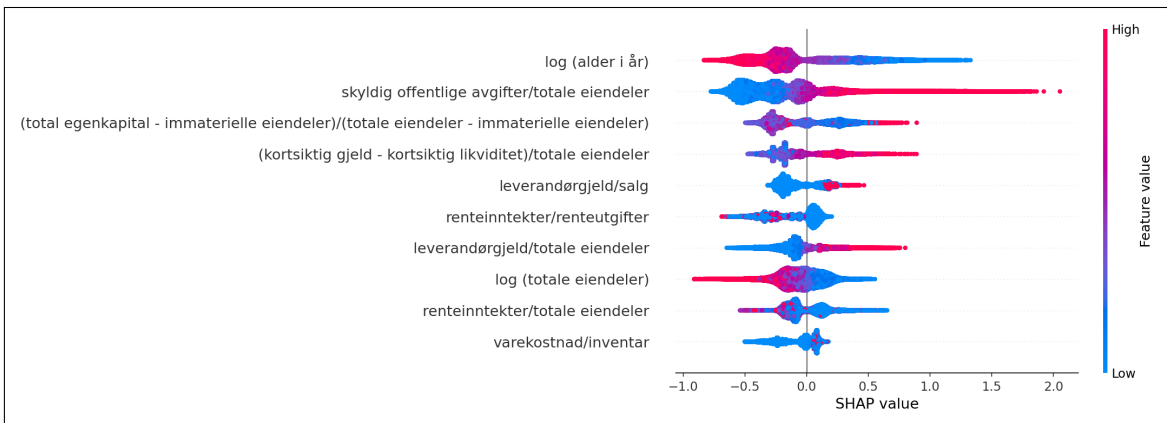
Figur 19: SHAP beeswarm plott for testår 2012. SHAP beeswarm plottet viser den globale variabelviktigheten for de ti viktigste variablene i modellen for det gitte teståret. Den viktigste variabelen er øverst, deretter følge de andre variablene i synkende rekkefølge. Hver prikk i plottet representerer én observasjon. Fargen på prikkene viser hvorvidt verdien for variabelen er lav (blå) eller høy (rød). X-aksen representerer SHAP-verdien til observasjonen.



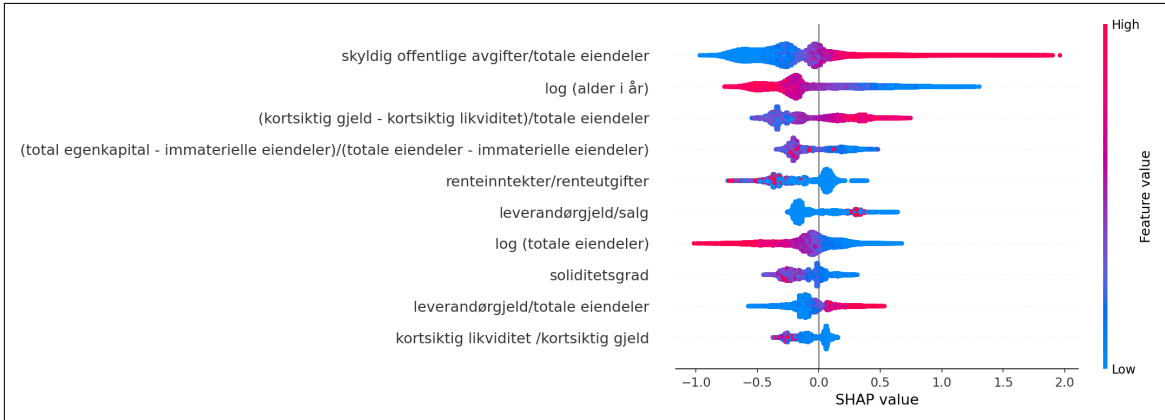
Figur 20: SHAP beeswarm plott for testår 2013. SHAP beeswarm plottet viser den globale variabelviktigheten for de ti viktigste variablene i modellen for det gitte teståret. Den viktigste variabelen er øverst, deretter følge de andre variablene i synkende rekkefølge. Hver prikk i plottet representerer én observasjon. Fargen på prikkene viser hvorvidt verdien for variabelen er lav (blå) eller høy (rød). X-aksen representerer SHAP-verdien til observasjonen.



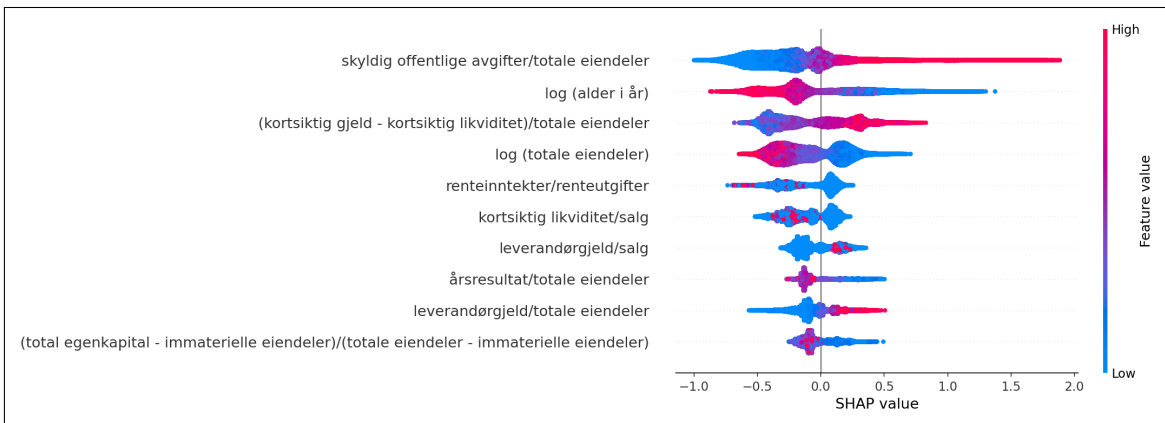
Figur 21: SHAP beeswarm plott for testår 2014. SHAP beeswarm plottet viser den globale variabelviktigheten for de ti viktigste variablene i modellen for det gitte teståret. Den viktigste variabelen er øverst, deretter følge de andre variablene i synkende rekkefølge. Hver prikk i plottet representerer én observasjon. Fargen på prikkene viser hvorvidt verdien for variabelen er lav (blå) eller høy (rød). X-aksen representerer SHAP-verdien til observasjonen.



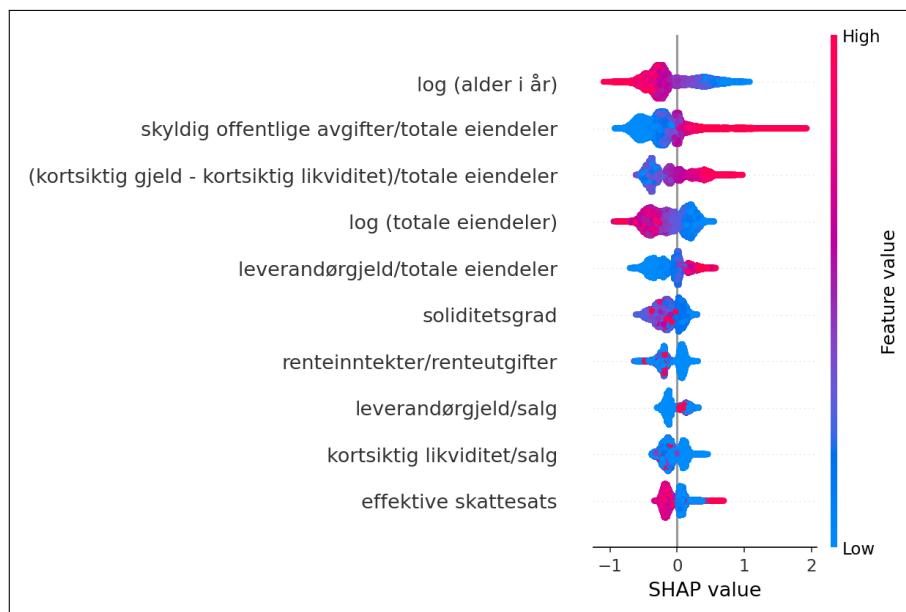
Figur 22: SHAP beeswarm plott for testår 2015. SHAP beeswarm plottet viser den globale variabelviktigheten for de ti viktigste variablene i modellen for det gitte teståret. Den viktigste variabelen er øverst, deretter følge de andre variablene i synkende rekkefølge. Hver prikk i plottet representerer én observasjon. Fargen på prikkene viser hvorvidt verdien for variabelen er lav (blå) eller høy (rød). X-aksen representerer SHAP-verdien til observasjonen.



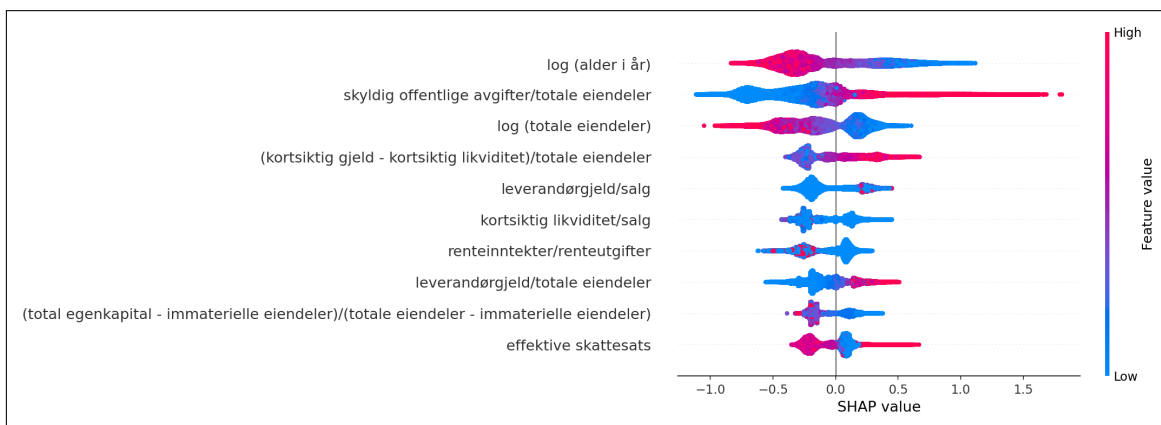
Figur 23: SHAP beeswarm plott for testår 2016. SHAP beeswarm plottet viser den globale variabelviktigheten for de ti viktigste variablene i modellen for det gitte teståret. Den viktigste variabelen er øverst, deretter følge de andre variablene i synkende rekkefølge. Hver prikk i plottet representerer én observasjon. Fargen på prikkene viser hvorvidt verdien for variabelen er lav (blå) eller høy (rød). X-aksen representerer SHAP-verdien til observasjonen.



Figur 24: SHAP beeswarm plott for testår 2017. SHAP beeswarm plottet viser den globale variabelviktigheten for de ti viktigste variablene i modellen for det gitte teståret. Den viktigste variabelen er øverst, deretter følge de andre variablene i synkende rekkefølge. Hver prikk i plottet representerer én observasjon. Fargen på prikkene viser hvorvidt verdien for variabelen er lav (blå) eller høy (rød). X-aksen representerer SHAP-verdien til observasjonen.



Figur 25: SHAP beeswarm plott for testår 2018. SHAP beeswarm plottet viser den globale variabelviktigheten for de ti viktigste variablene i modellen for det gitte teståret. Den viktigste variabelen er øverst, deretter følge de andre variablene i synkende rekkefølge. Hver prikk i plottet representerer én observasjon. Fargen på prikkene viser hvorvidt verdien for variabelen er lav (blå) eller høy (rød). X-aksen representerer SHAP-verdien til observasjonen.



Figur 26: SHAP beeswarm plott for testår 2019. SHAP beeswarm plottet viser den globale variabelviktigheten for de ti viktigste variablene i modellen for det gitte teståret. Den viktigste variabelen er øverst, deretter følge de andre variablene i synkende rekkefølge. Hver prikk i plottet representerer én observasjon. Fargen på prikkene viser hvorvidt verdien for variabelen er lav (blå) eller høy (rød). X-aksen representerer SHAP-verdien til observasjonen.

