Automatisering af indtjeningsestimater ved brug af Machine Learning

Kandidatafhandling Cand.merc. FIR

Forfattere:	**Studie-nummer:**
Frederik E.N. Christiansen	109694
Oliver Dollerup	110778

Antal anslag / Antal normalsider:	**211.415 / 92,92**
Vejleder:	Henning Skov Jensen
Afleveringsdato:	17 Maj 2021

Abstract

Despite evidence of their herding behavior and optimism bias, equity analysts have been acknowledged for providing the best proxy for future corporate earnings for many decades. Previous studies have documented equity analysts' superior forecast accuracy compared to statistical models throughout the 1980s and 1990s. The past decades have shown significant enhancements in CPU efficiency and computational power, enabling the use of big data and artificial intelligence. While the use of machine learning has revolutionized areas in the financial industry ranging from algorithmic trading to risk management and fraud detection, the application of machine learning has not been explored in the context of corporate earnings forecasts. This study leverages state-of-the-art machine learning algorithms to explore the potential of machine learning as a standalone estimator for future corporate earnings per share. By implementing a machine learning algorithm based on accounting figures and stock information we compare the forecast accuracy to those of equity analysts at horizons of one, two, and four quarters to evaluate the applicability of machine learning. We find that the Catboost algorithm produces more accurate earnings per share forecasts than equity analysts on average across all horizons, which is statistically significant on horizons of two and four quarters. In addition, we find that the Catboost algorithm forecasts are more accurate than equity analysts' when there is a low analyst coverage and a high forecast dispersion among analysts as well as when the market capitalization is small. The results of this study provide preliminary support for the potential to use machine learning as a standalone estimator for future corporate earnings.

Indhold

1 Indledning 1
1.1 Afgrænsning . . . 2
1.1.1 Data . . . 2
1.1.2 Inputvariable . . . 3

2 Litterature Review 4
2.1 Analytiker performance generelt . . . 4
2.1.1 Faktorer der påvirker analytikeres performance . . . 4
2.1.2 Incitamenter . . . 6
2.1.3 Indflydelse på markedet . . . 8
2.2 Automatisering af indtjeningsforecast . . . 9
2.2.1 Statistiske modeller . . . 9
2.2.2 Multivariate tidsserie forecast . . . 11
2.2.3 Machine Learning . . . 12

3 Metode 15
3.1 Data . . . 15
3.1.1 Beskrivelse af datagrundlaget . . . 15
3.1.2 Datakilder . . . 19
3.2 Data cleaning . . . 21
3.2.1 Fase 1 . . . 22
3.2.2 Fase 2 . . . 23
3.2.3 Fase 3 . . . 23
3.2.4 Fase 4 . . . 23
3.2.5 Fase 5 . . . 24
3.2.6 Fase 6 . . . 24
3.2.7 Fase 7 (Imputing) . . . 25
3.2.8 Fase 8 (Scaling) . . . 26
3.2.9 Behandling af outliers . . . 26

4 Teoretisk grundlag 28
4.1 Supervised Learning . . . 28
4.1.1 Generaliserbarhed . . . 29
4.2 Supervised machine learning algoritmer . . . 30
4.2.1 Lineære modeller . . . 30
4.2.2 Beslutningstræer . . . 31
4.2.3 Ensemble learning . . . 32
4.2.4 Support Vector Machine . . . 35

4.3 Evaluering af model performance . . . 36
4.3.1 Performance på samme datagrundlag . . . 36
4.3.2 Performance ved forskelligt datagrundlag . . . 37

5 Modellering 39
5.1 DataRobot . . . 39
5.2 Opdeling af data i trænings- og testsæt . . . 40
5.3 Cross validation . . . 40
5.4 Feature Engineering . . . 43
5.5 Feature selektion . . . 43
5.6 Hyperparameter tuning . . . 44

6 Resultater og analyse 46
6.1 Generel model resultater . . . 47
6.1.1 Forecast horisont på 1 kvartal . . . 47
6.1.2 Forecast horisont på 2 kvartaler . . . 48
6.1.3 Forecast horisont på 4 kvartaler . . . 48
6.1.4 Feature importance . . . 49
6.2 Industri . . . 51
6.2.1 Industrispecifikke machine learning modeller . . . 51
6.2.2 Model vs. analytikere . . . 53
6.3 Market cap . . . 57
6.4 Analytikerdækning og spredning i estimater . . . 61
6.4.1 Analytikerdækning . . . 61
6.4.2 Spredning i analytikerestimater . . . 62
6.5 Robusthed af resultater . . . 63
6.5.1 Corona krisen . . . 63
6.5.2 Længde af træningsdata . . . 65
6.5.3 Rolling træning test . . . 66

7 Diskussion 67
7.1 Begræsninger . . . 67
7.2 Refleksion . . . 68
7.3 Relater til tidligerelitteratur . . . 68

8 Konklusion og forlag til yderligere research 70
8.1 Konklusion . . . 70
8.2 Forslag til yderligere research . . . 71

A Inputvariable 72
A.1 Data variable før cleaning . . . 72
A.2 Baseline inputvariable . . . 74
A.3 Endelige inputvariable . . . 75

B SIC koder 77

C Analyse 78
C.1 Marketcap uden Healthcare 78

D Python 79

Bibliografi 80

Figurer

3.1 Forecast af EPS med machine learning - tidslinje 16
3.2 Antal observationer i datagrundlaget 17
3.3 Observationer fordelt på industrier 18
3.4 Data cleaning proces 22
3.5 Imputede inputvariable 25
3.6 Fordeling af realiseret EPS 27

4.1 Kompleksitet Vs. forecast præcision 29
4.2 Lineær regression eksempel 30
4.3 Beslutningstræ eksempel 31
4.4 Bagging metoden 33
4.5 Boosting metoden 34
4.6 Stacking metoden 34
4.7 Support Vector Regression (SVR) eksempel 35

5.1 Rolling origin cross validation eksempel 41
5.2 Cross validation Datarobot 42
5.3 Hyperparameter tuning - optimerings historik 45

6.1 Feature importance 50
6.2 Fordeling af observation på virksomhedsstørrelse 58
6.3 Outperformance før/under corona 64
6.4 Robusthedstjek af ændringer i længden af træningsdata 65
6.5 Robusthedstjek af rolling træning/test 66

Tabeller

5.1 Opdeling af datasæt 41
5.2 Catboost hyperparametre 45

6.1 Generel forecast performance 47
6.2 Industrispecifikke modeller Vs. generel model 52
6.3 Forecast performance opgjort pr. industri, $h = 1$ 54
6.4 Forecast performance opgjort pr. industri, $h = 2$ 55
6.5 Forecast performance opgjort pr. industri, $h = 4$ 56
6.6 Forecast performance opgjort på market cap, $h = 1$ 58
6.7 Forecast performance opgjort på market cap, $h = 2$ 59
6.8 Forecast performance opgjort på market cap, $h = 4$ 60
6.9 Modellens outperformance af analytikere opgjort på analytikerdækning 61
6.10 Modellens outperformance af analytikere opgjort på spredning i analytiker estimater 63

A.1 Datavariable før cleaning 72
A.2 Inputvariable i baseline model 74
A.3 Inputvariable i den endelige model 75
A.3 Inputvariable i den endelige model 76

B.1 Beskrivelse af industrier 77

C.1 Forecast performance market cap renset for *Healthcare* 78

1 Indledning

Præcise og korrekte indtjeningsestimater har stor påvirkning på adskillige områder inden for akademia, såvel som interessenter inden for den finansielle sektor. Indtjeningsestimater anvendes som primært input i alt fra værdiansættelser, aktivallokering, samt diverse finansielle og regnskabsmæssige nøgletal, herunder implied cost of capital og fremtidig price-earnings ratio (Schreiner & Spremann, 2007). Siden 1980'erne har der blandt akademikere og praktikere, eksisteret bred enighed om, at aktieanalytikeres indtjeningsestimater udgør den bedste proxy for selskabers fremtidige indtjening, grundet analytikeres overlegne forecast præcision sammenlignet med statiske modeller (Brown et al., 1987). I litteraturen er flere problematikker forbundet med aktieanalytikeres indtjeningsestimater identificeret. Litteraturen historisk kunne dokumentere empirisk evidens for systematiske fejl i analytikeres indtjeningsestimater i form af herding og optimisme bias (Beckers et al., 2004).

Ydermere er der de seneste år sket et fald i analytikerdækningen som følge af regulatoriske ændringer i aktieanalytikeres arbejdsvilkår, der blandt andet har begrænset adgang til selskabers ledelse. Machine learning kan potentielt løse nogle af disse udfordringer, og dermed give en bedre proxy for selskabers fremtidige indtjening, da brugen af machine learning har vist sig at revolutionere utallige andre områder i den finansielle sektor. I den finansielle sektor benyttes machine learning blandt andet i forbindelse med identificering af regnskabssvindel og risikostyring samt inden for algoritmisk trading, hvor hedgefonde såsom Walnut Investments og Renaissance Technologies har haft stor succes med implementeringen af machine learning (Faggella, 2020).

Set i lyset af machine learnings succesfulde implementering i den finansielle sektor er det overraskende at forskningen indenfor anvendelsen af machine learning i forbindelse med automatisering af indtjeningsestimer er yderst begrænset. I litteraturen udgør Elend et al. (2020) det eneste studie, der undersøger machine learning modellers anvendelighed i forbindelse med estimering af selskabers fremtidig indtjening. Elend et al. (2020) resultater var banebrydende, da disse indikerede at en machine learning model, på baggrund af regnskabstal og aktiedata, opnåede mere præcise indtjeningsestimater end aktieanalytikere. Elend et al. (2020) banebrydende resultater understreger machine learnings potentiale for at automatisere indtjeningsestimater, men berør kun overfladen med sine præliminære resultater.

Afhandlingen ønsker hertil at skabe et mere nuanceret indblik i anvendeligheden af machine learning algoritmer, baseret på regnskabs- og aktieinformations, som proxy for selskabers fremtidige indtjening, hvortil følgende problemstilling er udarbejdet:

> *I hvilket omfang kan en machine learning model baseret på regnskabstal og aktiedata anvendes til at estimere selskabers fremtidige indtjening per aktie?*

Med henblik på at strukturere samt sikre en fyldestgørende besvarelse af den fremstillede problemstilling, vil følgende arbejdsspørgsmål løbende besvares i afhandlingen:

- Hvilken indflydelse har forecast horisonten på machine learning modellens anvendelighed?
- Er der industrier, hvor machine learning modellen har større anvendelighed end andre, samt er det mest hensigtsmæssigt at implementere industrispecifikke modeller frem for en generel model?
- Hvilken indflydelse har selskabers størrelse samt analytikerdækning på machine learning modellens anvendelighed?

Formålet med afhandlingen er at give akademikere og praktikere en bedre forståelse af machine learnings anvendelighed samt et indblik i, hvornår det er mest optimalt at benytte machine learning baserede forecast fremfor analytikerestimater, som proxy for selskabers fremtidige indtjening. Afhandlingens primære bidrag til litteraturen vil hertil være at skabe et mere nuanceret billede af, i hvilket omfang, samt i hvilke sammenhænge machine learning baserede forecast udgør en hensigtsmæssig proxy for selskabers fremtidige indtjening. Dette undersøges gennem en af dekomponering af modellens resultaterne fordelt på forskellige industrier, virksomhedstørrelser samt analytikerdækning.

1.1 Afgrænsning

Da der i afhandlingen lægges stor vægt på at danne en fundamental forståelsesramme for machine learning algoritmer og anvendelsen heraf, qua afhandlingens målgruppe, er matematiske udledninger begrænset. Ydermere inddrager afhandlingen ikke implementeringsmæssige og praktiske implikationer ved anvendelsen machine learning, herunder dataindsamling og automatisering heraf samt regulatoriske forhold.

1.1.1 Data

Afhandlingens datagrundlag vil udelukkende bestå af amerikanske selskaber, da data vedrørende amerikanske selskaber ofte er mere fyldestgørende end eksempelvis data vedrørende europæiske selskaber. Størrelsen samt kompletheden af data er yderst vigtigt i forbindelse med implementering af machine learning modeller, hvor store mængder data ofte er et succeskriterie for en god model. Endvidere vil afhandlingens datagrundlag alene bestå af børsnoterede selskaber, da private virksomheder ikke er forpligtet til at offentliggøre regnskabstal i USA, hvorfor denne afgrænsning er en naturlig konsekvens af den gældende regulering.

Til vurderingen af machine learning modellens anvendelighed vil modellens forecast præcisionen sammenlignes med analytikeres konsensus estimat, der udgør et gennemsnit af alle analytikeres estimater for et givent selskab på et givne tidspunkt. Årsagen til valget af konsensusestimatet fremfor individuelle analytikerestimater er, at konsensusestimat afspejler det estimat, der anvendes af akademikere og praktikere som proxy for selskabers fremtidige indtjening.

Datagrundlaget i afhandlingen afgrænses til data fra perioden 1995-2020. Den valgte tidsperiode for datagrundlaget tager højde for ændringer i økonomiske konjunkturer, såvel som strukturelle ændringer i forretningsmodeller samt omkostningsstrukturer og minimerer dermed time period bias (CFI Institute, 2021). Datagrundlaget giver dermed et repræsentativt grundlag for analysen samt for generaliserbarheden i afhandlingens konklusioner.

1.1.2 Inputvariable

Valget af inputvariable i machine learning modellen udgør et essentielt element i en modelleringsprocessen, hvortil afhandlingens fokus på at undersøge machine learnings anvendelighed som proxy i akademia og praksis kræver, at inputdata er let lettilgængelig. Det er hertil fundet hensigtsmæssigt at begrænse inputvariable til fundamentale regnskabstal, da det i studierne af Elend et al. (2020) og Ball & Ghysels (2017) blev konstateret, at netop disse inputvariable havde en høj forklaringskraft for selskabers fremtidig indtjening. Ydermere benyttes aktiedata i form af aktieafkast og aktiekurser ligeledes som inputvariable, da W.Beaver et al. (1980), Collins et al. (1987) og Freeman (1987) dokumenterede at aktieafkast opfanger information om selskabers fremtidige indtjening.

2 Litterature Review

Formålet med kapitlet er at kortlægge relevant litteratur samt at skabe et overblik over afhandlingens primære fokusområder. I kapitlet vil læseren præsenteres for forskellige aspekter af analytikeres forecast præcision, beskrevet i litteraturen. Ydermere vil forskellige modeller anvendt i litteraturen i forbindelse med automatisering af selskabers indtjeningsestimater, samt resultaterne heraf præsenteres. Kapitlet vil endvidere gøre det muligt at evaluere, hvordan afhandlingen bidrager til den eksisterende litteratur og derigennem skabe fundamentet for en diskussion af afhandlingens resultater.

Key takeaways: Litteraturen har siden 1980erne fastslået, at det ikke var muligt for statistiske modeller at opnå en højere forecast præcision end analytikere. Nyere studier af Ball & Ghysels (2017) og Elend al. (2020) viste dog banebrydende resultater, da disse ved anvendelse af avancerede statistiske modeller og machine learning formåede at outperforme analytikere ved estimering af amerikanske selskabers EPS.

2.1 Analytiker performance generelt

Analytikeres forecast præcision er beskrevet omfattende i litteraturen igennem årene, og har siden årtusindeskiftet haft fokus på at identificere faktorer, der påvirker analytikeres performance. Analytikeres forecast præcision har været i fokus, grundet både praktikere og akademikeres anvendelse af disse i forbindelse med værdiansættelser, aktivallokering, samt diverse finansielle og regnskabsmæssige nøgletal. Litteraturen indenfor analytikeres forecast præcision kan hovedsageligt inddeles i de tre fokusområder; faktorer der påvirker analytikeres estimater, analytikeres incitamenter samt analytikeres indvirkning på de finansielle markeder. På baggrund heraf vil afsnittet inddeles i disse hovedområder og vil hertil beskrive relevante resultater fremstillet i litteraturen.

2.1.1 Faktorer der påvirker analytikeres performance

Eksisterende litteratur omhandlende faktorer, der påvirker analytikeres forecast præcision, anskuer området fra forskellige perspektiver. Visse studier identificerer virksomheds- og analytikerspecifikke karakteristika, der påvirker analytikeres forecast præcision, mens andre studier fokuserer på eksterne faktorer såsom kvaliteten af regnskaber og ændring i regnskabspraksis.

Virksomheds- og analytikerspecifikke faktorer

Et stort omfang af studier identificerede, at forskellige analytikerspecifikke karakteristika havde indflydelse på analytikeres forecast præcision (Athavale et al., 2013; Barniv et al., 2005; Clement, 1999; Jacob et al., 1999; Myring & Wrege, 2009). Jacob et al. (1999) undersøgte hvorvidt

analytikerspecifikke karakteristika herunder; forecast horisont, ændring i antal analytikere, analytikeres erfaring, antal selskaber analytikeren dækker, størrelse på broker, kunne forklare variation i analytikeres forecast præcision. Jacob et al. (1999) konstaterede hertil en stærk relation mellem de analytikerspecifikke variable og analytikeres forecast præcision. Clement (1999) undersøgte ligeledes indflydelsen af analytikerspecifikke karakteristika på analytikeres forecast præcision, og valgte hertil at inkorporere yderligere variable for analytikeres erfaring. Resultaterne fremført af Clement (1999) var konsistente med konklusionerne fremstillet af Jacob et al. (1999). Ifølge Maines et al. (1997) har analytikeres erfaring en essentiel betydning for analytikeres forecast præcision, hvilket endvidere kommer til udtryk i Maines et al. (1997) resultater, der vidnede om en sammenhæng mellem analytikeres forecast præcision og analytikeres erfaring. Maines et al. (1997) konkluderer i studiet, at erfarne analytikeres forecast estimater bærer præg af markant større præcision end forecast fra uerfarne analytikere. Mikhail et al. (1997) konkluderer hertil, at analytikeres erfaring med individuelle virksomheder øger analytikerens forecast præcision. Ifølge Mikhail et al. (1997) vil analytikere, der har indgående kendskab til individuelle virksomheder have et mere solidt fundament at basere forecast på, hvilket reducerer forecast fejlen i analytikeres estimater.

Metoderne udviklet af Jacob et al. (1999) og Clement (1999) er inkorporeret i mange nyere studier, herunder af Bae et al. (2008), der undersøgte forskellen mellem henholdsvis indenlandske og udenlandske analytikeres forecast præcision. Ydermere benytter Jacob et al. (2008) delkomponenter af Jacob et al. (1999) og Clement (1999) metode til at undersøge, hvorvidt analytikere med tilknytning til investeringsbanker opnår en højere forecast præcision. Jacob et al. (2008) konkluderer hertil, at analytikere med tilknytning til investeringsbanker gennemsnitligt har en højere præcision end analytikere uden tilknytning. Udover virksomheds- og analytikerspecifikke karakteristika, har tidligere studier ligeledes identificeret, at de makroøkonomiske forhold der omgiver selskaber, har indflydelse på analytikeres forecast præcision. Ifølge Chopra (1998) steg præcisionen i analytikeres indtjeningsestimater, såfremt der forekom en høj økonomisk vækst. Modsat konstaterede Chopra (1998) et fald i forecast præcisionen, når der ikke forekom en høj økonomisk vækst. Hope & Kang (2005) undersøgte ligeledes hvordan makroøkonomisk usikkerhed påvirker analytikeres præcision på internationalt plan. Ved at benytte inflation og valutakurs volatilitet, som mål for den makroøkonomiske usikkerhed, dokumenterede Hope & Kang (2005) evidens for, at analytikeres forecast præcision falder i takt med, at den makroøkonomiske usikkerhed stiger. Ydermere konstaterede Hope & Kang (2005), at den negative relation mellem analytikeres forecast præcision og makroøkonomisk usikkerhed var stærkest for udviklingslande.

Regnskabsrelaterede faktorer

Analytikeres primære arbejde består i at evaluere selskabers performance, hvilket kræver information fra mange forskellige informationskilder. Detaljerede finansielle opgørelser såsom resultatopgørelse, balance og pengestrømsopgørelser, udgør hertil et væsentligt grundlag for analytikeres indtjenings estimater (Brown et al., 2014). I litteraturen har adskillelige forskere undersøgt hvorvidt der fremgår en sammenhæng mellem forecast præcisionen blandt analytikere og kvaliteten af underliggende finansielle opgørelser (Dhaliwal et al., 2012; Holland, 1998; Kwag & Small, 2007; Lang & Lundholm, 1996; Tong, 2007).

Ifølge Holland (1998), fremgik der en højere forecast præcision i analytikeres estimater, når der forelå detaljeret information i forlængelse af finansielle opgørelser, i form af beskrivende og forklarende

noter. Ydermere udgjorde selskaber regnskabspraksis ligeledes en væsentlig faktor for analytikeres estimater. Hope (2003), der undersøgte sammenhængen mellem analytikeres præcision i indtjenings estimater og detaljeringsgraden i årsrapporter i mere end 22 lande, kunne konstatere at analytikeres forecast præcision var højere, i tilfælde hvor selskabernes regnskabspraksis fremgår eksplicit. Ydermere har studier vist, at detaljeringsgraden af ikke-finansielt data ligeledes har indflydelse på analytikeres estimater. På baggrund af en analyse af ikke-finansielt data, i form af CSR-rapporter fra mere end 31 lande, konkluderede Dhaliwal et al. (2012), at udgivelse af CSR-rapporter øgede analytikeres forecast præcision. Ydermere kunne Yu (2010), der undersøgte antallet af corporate governance disclosures i årsrapporten indflydelse på analytikernes forecast præcision, konstatere at analytikere foretager mere præcise indtjeningsestimater såfremt selskaberne offentliggør mere omfattende corporate governance disclosures. Kwag & Small (2007), der ligeledes undersøgte ikke-finansiel data indflydelse på analytikeres estimater, understøttede konklusionerne fremstillet af Dhaliwal et al. (2012) og Yu (2010), da de kunne konkludere at indførslen af Regulation Fair Disclosure (FD)[1] medførte et væsentligt fald i analytikeres forecast præcision. Modsat Kwag & Small (2007) konklusioner konkluderede Tong (2007), at internationale regulativer med fokus på at øge transparens, herunder FD. ikke havde en signifikant indflydelse på analytikernes forecast præcision.

Kvaliteten af rapporterede finansielle opgørelser, kan i mange tilfælde forklares af revisionskvaliteten. Endvidere kan revisionskvaliteten anskues på baggrund af forskellige variable såsom størrelsen på revisionsselskabet samt hvorvidt revisionsselskabet har specialiseret sig inden for en given industri. Behn et al. (2007) undersøgte hvorvidt kvaliteten af revision har indflydelse på analytikeres evne til at estimere fremtidig indtjening og fandt hertil evidens for, at analytikeres estimater havde højere præcision, når revisionen er udført af enten Big Four (EY, Deloitte, KPMG, PwC) eller en industri specifik revisor. Resultaterne indikerede dermed, at en bedre analytiker performance kan relateres til revision af høj kvalitet enten i form af Big Four eller en industrispecialist. Xie et al. (2012) kunne ligeledes bekræfte den positive relation mellem analytikeres forecast præcision og kvaliteten af finansiel rapportering fremstillet af Behn et al. (2007).

2.1.2 Incitamenter

Analytikere har gennem årene været fokuspunkt for kritik og debat grundet kvaliteten af deres estimater. Akademisk research i mange år kunne dokumentere systematiske fejl i analytikeres indtjeningsforecast, der hovedsageligt kan opdeles i kategorierne *Optimsme* og *Herding* (Beckers et al., 2004)

Optimisme bias

En af de mest udbredte, konsistente og robuste kognitive bias inden for psykologi og behaviorial economics er optimisme bias, der er identificeret blandt mennesker på tvært af lande, kulturer og uddannelsesniveauer (Sharot, 2011). Optimisme bias handler i sin simpleste form om, at vi mennesker overestimater positive begivenheder og underestimerer negative begivenheder (Sharot, 2011).

[1] FD er en regel vedtaget af Securities and Exchange Commission (SEC) i et forsøg på at forhindre offentlige virksomheders selektive offentliggørelser til professionelle på markedet og visse aktionærer.

Optimisme bias blandt aktieanalytikere er undersøgt på adskillelige markeder de sidste 25-30 år. Blandt andet undersøgte Dreman & Berry (1995) analytikeres indtjeningsestimater på det amerikanske marked i perioden 1974 til 1991, og identificerede hertil en gennemsnitlig afvigelse på 20% i forhold til det realiserede EPS. Dreman and Berry (1995) konkluderede hertil, at analytikere udviste optimisme bias i deres EPS forecast i hele perioden, samt at der ikke forekom en signifikant forskel i optimisme biaset på tværs af industrier eller økonomiske cyklusser. Optimisme bias blandt analytikere er yderligere fundet statistisk signifikant på andre markeder, herunder Storbritannien (Capstaff et al., 1995; De Bondt & Forbes*, 1999), Tyskland (Capstaff & Rees, 1998), Europa (Beckers et al., 2004; Capstaff et al., 2001) og senest Frankrig (Galanti & Vaubourg, 2017). Beckers et al. (2004) adskilte sig fra tidligere research ved at identificere faktorer, der påvirkede bade Optimisme bias og analytikeres forecast præcision mere generelt. Beckers et al. (2004) kunne hertil konstatere, at optimisme bias og høje forecast fejl var mere udbredt blandt selskaber med en højere historisk aktie volatilitet samt med en høj spredning af analytiker forecast. Ydermere identificerede Beckers et al. (2004) en væsentlig forskel i optimisme bias på tværs af industrier, herunder blev forbruger-, sundhed-, forsynings- og transportsektorerne identificeret som de brancher med lavest forecast fejl, og den laveste grad af optimisme bias. Industriforskellen fremført af Beckers et al. (2004) er konsistente med resultaterne fremstillet af Capstaff et al. (2001), der ligeledes konstaterede industriforskelle. Beckers et al (2004) bekræftede ydermere tidligere research på området (Capstaff et al., 2001; Capstaff & Rees, 1998; De Bondt & Forbes*, 1999), der viste en højre forecast fejl samt optimisme bias på længere forecast horisonter.

Litteraturen har længe forsøgt at klarlægge årsagerne bag optimisme bias. En af de primære årsager, identificeret i litteraturen er interessekonflikten mellem Equity research og Investment banking afdelingen (Arand & Kerl, 2015; Devos, 2014). Interesse konflikten opstår ved, at såfremt analytikere udsteder positive rapporter og analyser omhandlende et givent selskab, vil det alt andet lige være nemmere for Investment banking afdelingen at skabe eller opretholde kundeforholdet (Brien et al., 2005; Dechow et al., 2000; McKnight et al., 2010; Michaely & Womack, 1999). Interessekonflikt kommer blandt andet til udtryk SEC forliget kendt under navnet "Global Research Analyst Settlement"[2].

Ydermere er interessekonflikten for analytikere ansat af brokers i litteraturen identificeret som en årsag til optimisme bias (Carleton et al., 1998; Jackson, 2005; Mehran & Stulz, 2007). Optimistiske forecast blandt analytikere vil alt andet lige generere flere købsordrer, og dermed øge indtjeningen for brokers i form af kurtage.

Herding

Herding er et veletableret socialt fænomen, der er bredt diskuteret blandt akademikere, og akademisk forskning indenfor området har resulteret i en lang række teorier. Herding blandt analytikere siges at forekomme, såfremt analytikere i forbindelse med forecast vælger at se bort fra egne estimater, og i stedet forecaste mod konsensus fra andre analytikere (Bernhardt et al., 2006).

Klassisk økonomisk teori fastslår, at investeringsbeslutninger baserer sig på rationelle forventninger med udgangspunkt i al offentlig tilgængelig information. Keynes (1936) stillede dog spørgsmålstegn til denne opfattelse, da investorer ifølge ham har en tendens til at agere på baggrund af egen information og interesser, grundet en frygt for, at modstridende adfærd vil skade deres omdømme.

[2] Forlig med ti af de største investment banks.

Scharfstein og Stein (1990) byggede videre på Keynes (1936) opfattelse, ved at undersøge herding adfærd i en forretningsmæssig sammenhæng, og kunne hertil konkludere at herding forekommer i forskellige sammenhænge, grundet managers naturlige forsøg på at forbedre deres omdømme (Se også: Rajan, 2006). Ydermere konkluderede Scharfstein og Stein (1990) at incitamentet til herding er størst, når managers karrieremuligheder uden for virksomheden er uattraktive, samt når aflønning baseres på absolutte frem for relative performance mål.

Hong et al. (2000) valgte at undersøge herding blandt analytikere, samt de implicitte incitamenter analytikere står overfor. Hong et al. (2000) konkluderede hertil at henholdsvis uerfarne og erfarne analytikere har forskellige incitamenter. Uerfarne analytikere straffes i højere grad for utilstrækkelig forecast performance og forecast der afviger fra konsensus. I forlængelse heraf indikerede resultaterne fra studiet, at uerfarne analytikere udviser en højere grad af herding, i form af, at forecaste tættere på konsensus end erfarne analytikere. Stickel (1990) og Graham (1999) konstaterede en tendens til, at herding steg i takt med antallet af estimater, der lå tæt på konsensusestimatet, samt i forbindelse med en lav forecast præcision i analytikeres tidligere estimater. Graham (1999) forskning baserede sig på Scharfstein og Stein (1990) model, og resultaterne heraf indikerede, at såfremt analytikere havde et godt omdømme eller en lav evne, var analytikere tilbøjelige til herding. Bikhchandani et al. (1992) argumenterede hertil for, at det i visse tilfælde er mest optimalt for analytikere, at se bort fra privat information og følge andres anbefalinger, da disse besidder vigtige informationer. Welch (2000) resultater indikerede ligeledes herding blandt analytikere, da det fremgik heraf, at analytikeres køb og salg anbefalinger, havde en signifikant indflydelse på de to efterfølgende analytikeres anbefalinger. Welch (2000) resultater var dog modstridende med Bikhchandani et al. (1992) argumentation, da resultaterne viste, at der ingen sammenhæng var mellem anvendelsen af konsensusestimater og præcisionen heraf. Welch (2000) resultater indikerede dermed, at indflydelsen af konsensus estimater ikke opstod i forsøget på at opnå yderligere information, der kunne forbedre analytikerens egne anbefalinger.

Nyere research på området præsenterer empiriske beviser for herding adfærd blandt analytikere, samt bekræfter væsentlige hovedpointer fra forhenværende studier, hvilket vidner om, at herding adfærd stadig forekommer blandt analytikere i nyere tid (Keskek et al., 2014; Leece & White, 2017)

2.1.3 Indflydelse på markedet

I modsætning til tidligere nævnte studier, der fokuserer på analytikeres adfærd samt faktorer der påvirker præcisionen i indtjeningsestimater, har andre studier forsøgt at adressere analytikeres indflydelse på de finansielle markeder (Brown & Hugon, 2009; T. C.-T. Hou et al., 2014; Jung et al., 2019). Litteraturen for analytikeres indtjenings estimater har vist en konsistent positiv markedsreaktion, når analytikere foretager ændringer i indtjenings estimater. I USA dokumenterer studier af Jung et al. (2019) samt Brown & Hugon (2009) en positiv relation mellem analytikeres indtjeningsestimater og afkast på aktiemarkedet. Ligeledes dokumenterer Bandyopadhyay et al. (1995) og Arand et al. (2015) tilsvarende resultater i henholdsvis Canada og resten af verden. Hou et al. (2014), der undersøgte selvsamme relation på det Australske marked, konkluderede at aktier med en opadgående ændring i indtjenings estimatet oplevede positive afkast, mens aktier med nedadgående ændringer havde negative afkast. Denne effekt var mest fremtrædende blandt aktier med høj informationsusikkerhed (T. C.-T. Hou et al., 2014).

Michel (2017), der undersøgte investorers reaktion på analytikeres forecast, identificerede at investorer overreagerer på forecast, der er identisk med minimum eller maksimum af ledelsens guidance spænd. Michel (2017) henviste hertil til, at overreaktion var drevet af overselvsikre investorer, der investerer på baggrund af informations usikkerhed. Yderligere studier på området konkluderer, at investorers reaktion på henholdsvis star-analytikere[3] og ikke star-analytikeres estimater ikke afviger (Kerl & Ohlert, 2015). Resultaterne indikerede dermed, at markedet ikke tager højde for analytikeres historiske evne til at estimere indtjening.

2.2 Automatisering af indtjeningsforecast

Forskningen inden for analytikeres indtjeningsestimater startede, da man inden for akademia ønskede at finde en hensigtsmæssig proxy for forventninger til virksomheders fremtidige indtjening, herunder særligt EPS (Bradshaw et al., 2012). Litteraturen har hertil beskæftiget sig med at automatisere indtjeningsestimater, med henblik på at opnå det bedste estimat for selskabers fremtidige EPS.

Fokusset på at udvikle en proxy for fremtidig EPS opstod som led i, dengang banebrydende viden om sammenhængen mellem regnskabstal og aktiekurser fremført Ball & Brown (1968). I 1970erne og 1980erne offentliggjorde adskillige studier resultater, der understøttede konsensus om, at analytikeres estimater var det bedste bud på selskabers fremtidige EPS, hvilket sidenhen blev den generelle opfattelse på området (Brown, Hagerman, et al., 1987, 1987; Brown & Rozeff, 1978; Collins & Hopwood, 1980; Fried & Givoly, 1982; O'brien, 1988).Litteraturen har siden 1980'erne og 1990'erne beskæftiget sig minimalt med indtjeningsforecast, hvilket Kothari (2001) beskriver således:

> I deliberately keep my remarks on the earnings' time-series properties short because I believe this literature is fast becoming extinct. The main reason is the easy availability of a better substitute: analysts' forecasts are available at a low cost in machine-readable form for a large fraction of publicly traded firms. (s.145)

Det seneste årti er der sket et paradigmeskifte inden for forskning af indtjeningsforecast, da introduktionen af MIDAS (Mixed frequency), big data og machine learning, har muliggjort udviklingen af nye tidserie modeller. I det følgende afsnit vil litteraturen inden for indtjeningsforecast hertil inddeles i de tre overordnede forecast metoder anvendt på området, herunder statistiske modeller og machine learning.

2.2.1 Statistiske modeller

Adskillelige studier har undersøgt tidsserie egenskaberne ved selskabers årlige indtjening og har hertil konstateret, at selskabers indtjening approksimativt følger en random walk udvikling, hvilket blev den generelle opfattelse i slutning af 1970'erne Little, 1962; R. Ball & Watts, 1972; Brown, Hagerman, et al., 1987). Mange studier tager udgangspunkt i denne generelle opfattelse, og sammenligner hertil forecast præcisionen mellem tidsserie forecast og analytiker forecast. Disse studier kan overordnet set inddeles i to kategorier, henholdsvis studier der ønsker at klarlægge

[3] Henviser til højt rankede analytikere i Wall Street Journal

hvorvidt analytikeres forecast præcision var overlegen i forhold til tidsserie forecast og studier, der undersøger årsagen til analytikeres tilsyneladende overlegenhed.

Fried & Givoly (1982) undersøgte, hvorvidt analytikeres forecast var overlegne i forhold til tidsseriemodeller, og argumenterede hertil for, at analytikeres outperformance af tidsseriemodeller skyldtes analytikeres informations fordel. Fried & Givoly (1982) resultater viste endvidere, at analytikeres forecast fejl var 16,4%, og dermed outperformede en random walk model med en forecast fejl på 19,3%, hvilket var konsistente resultaterne fremstillet i lignende studier. Blandt andet fremgik det af Collins & Hopwood (1980) resultater, at analytikeres gennemsnitlige forecast fejl var 31,7% sammenlignet med 32,9% for deres konstruerede tidsseriemodel. Modsat tidligere studier valgte Brown et al. (1987) at undersøge årsagen til analytikernes outperformance af statistiske modeller. Brown et al. (1987) resultater viste endvidere en positiv sammenhæng mellem virksomhedsstørrelsen og analytikeres outperformance, og ydermere en negativ sammenhæng mellem analytiker performance og spredningen i analytikerestimaterne. Brown et al. (1987) og andre lignende studier, herunder Kross et al. (1990) og Lys and Soo (1995) konstaterede ligeledes en negativ korrelation mellem analytikeres outperformance af statistiske modeller og længden af forecast horisonten. Brown et al. (1987) konkluderede i sit studie at analytikeres outperformance primært skyldtes, analytikeres adgang til mere information, samt et timings element. Ydermere bekræftede Kross et al. (1990) og Lys and Soo (1995) timing effekten, der blandt andet fremgår ved at statistiske modeller baseret på regnskabstal ikke har adgang til information mellem regnskabstallenes offentliggørelse og forecast tidspunktet, hvilket analytikere har. Ydermere argumenterede O'brien (1988) for, at analytikeres outperformance af statistiske modeller skyldtes analytikeres større adgang til information herunder industrispecifik- og makroøkonomiskinformation samt andre analytikeres forecast, og bekræftede dermed Fried & Givoly (1982).

De statistiske modeller anvendt i de fremstillede studier, der forsøgte at klarlægge hvorvidt analytikeres forecast præcision var overlegen, baserede sig hovedsageligt på Box-Jenkins (ARIMA) eller random walk modeller. Box Jenkins forsøger grundlæggende at forecaste fremtidige værdier af en tidsserie ved at benytte de tre komponenter; autoregressiv, differentiering og moving average (Box et al., 2015, s. 88–97). Forecasting af earnings ved brug af Box-Jenkins er endvidere undersøgt af Brown & Rozeff (1978), Collins and Hopwood (1980), Brown et al. (1987), Lys & Soo (1995). I modsætning til Box-Jenkins modeller tager random walk modeller udgangspunkt i den grundlæggende antagelse, at den næste værdi i en tidserie baserer sig på den forrige værdi kombineret med en tilfældighed (Nau, 2020, par. Random Walk Model). Random walk modeller benyttes blandt andet af Brown et al. (1987), Foster (1977), Lorek (1979), Bathke og Lorek (1984) og senest i forbindelse med Bradshaw et al. (2012) genundersøgelse af litteraturens resultater.

Bradshaw et al. (2012) genundersøgelse af litteraturen resultater, der indikerede at analytikere outperformede tidsseriemodeller, afslørede at tidligere studier primært baserede sig på mindre datagrundlag, bestående af store veletablerede og stabile virksomheder. I studiet konkluderede Bradshaw et al. (2012), at analytikeres forecast præcision konsekvent outperformede et random walk forecast ved en forecast horisont på mindre end et år. Det fremgik endvidere af Bradshaw et al. (2012) resultater, at analytikeres outperformance af random walk forecast var størst, med en forskel på 282 basis point, ved forecast foretaget en måned før offentliggørelse af regnskabet. Resultaterne indikerede dog, at analytikernes outperformance på kort sigt blev gradvist forringet for hver ekstra måned tillagt forecast horisonten. Ydermere indikerede resultater, at analytikeres forecast præcision var aftagende på længere horisonter, svarende til forecast horisonter på to til tre år, hvortil random walk forecast viste sig at opnå højere forecast præcision på visse forecast horisonter. Bradshaw et al.

(2012) kunne hertil konstatere, at et random walk forecast opnåede en markant bedre forecast præcision end analytikere, såfremt der var tale om små og nyopstartede selskaber samt i de tilfælde, hvor analytikere forecastede ekstreme ændringer i EPS. Bradshaw et al. (2012) konkluderede på baggrund af resultaterne, at investorer med fordel kan stole på analytikeres forecast ved en forecast horisont på ét år eller derunder. Bradshaw et al. (2012) konklusioner var dermed konsistente med konklusionerne fremstillet i litteraturen i 1970-1990erne, men modsvarede samtidig litteraturen i den forstand, at analytikeres forecast dominans sammenlignet med tidsserie modeller primært forekommer på kort sigt.

2.2.2 Multivariate tidsserie forecast

I litteraturen vælger hovedparten af studier at benytte regnskabstal som det primær inputdata i de automatiserede forecast modeller. Ou & Penman (1989) fandt hertil frem til 68 finansielle nøgletal, der kunne benyttes i en logistisk regression til at forudsige, hvorvidt en given virksomheds indtjening steg eller faldt ved næstkommende årsrapport. Ou & Penman (1989) benyttede de finansielle nøgletal i en model til at konstruere lange og korte porteføljer, der endvidere viste sig at generere positive afkast på en henholdsvis 12 og 24 måneders horisont i perioden 1973 til 1983. Holthausen & Larcker (1992) benyttede ligeledes finansielle nøgletal og baserede disse på Ou og Penman (1989) i en logistisk regression til modellering af merafkast 12 måneder efter annoncering af virksomhedernes indtjening. På baggrund af modellen påviste Holthausen & Larcker (1992) et merafkast i intervallet 4,3% til 9,5% i perioden 1978 til 1988. Holthausen & Larcker (1992) merafkast viste sig endvidere ikke at være signifikante, og de kunne dermed ikke replikere Ou & Penman (1989) resultater.

Modsat tidligere studier, der benyttede en trinvis regression til at udvælge inputvariable, baserede Lev & Thiagarajan (1993) udvælgelse af nøgletal og inputvariable sig på, hvad analytikere lagde vægt på i artikler herunder eksempelvis Wall Street Journal, Value Line, samt aktieanalyse rapporter. Lev & Thiagarajan (1993) anvendelse af nøgletal viste endvidere en gennemsnitligt R^2 værdi på omtrent 0,7 på baggrund af data fra 1980'erne. I et nyere studie fra 2017 benyttede Yan & Zheng (2017) data science til at konstruere 18.000 fundamentale ratios på baggrund af 240 regnskabsvariable og 67 finansielle nøgletal og resultaterne heraf indikerede, at en stor del af de konstrueret nøgletal havde en signifikant forklaringskraft for udviklingen i aktieafkastet. Ydermere konstaterede Yan & Zheng (2017), at evnen til at forudsige aktieafkastet var bedst blandt små og udækkede virksomheder. Yan & Zheng (2017) resultater understregede endvidere anvendeligheden af data science i forbindelse med finansielle- og regnskabsmæssige problemstillinger.

Inden for den finansielle sektor har adgang til højfrekvensdata muliggjort anvendelsen af nye og mere sofistikerede forecast modeller (Subrahmanyam, 2019). Et eksempel herpå er Mixed frequency regression (MIDAS) modellen, der forsøger at løse nogle af udfordringerne ved Box-Jenkins, herunder kravet om data på samme frekvens (R. T. Ball & Ghysels, 2017). MIDAS modellen blev introduceret i 2004, af Ghysels, Santa-Clara og Valkanov, der benyttede modellen til at analysere forholdet mellem aktieafkast og volatilitet (Ghysels et al., 2004, 2005). MIDAS modellen er sidenhen hyppigt anvendt af andre forskere, særligt i forbindelse med forecasting af makroøkonomiske variable (Clements et al., 2008; Andreou et al., 2013; Duarte et al., 2017; Zhang & Wang, 2019; Gunay et al., 2020). Ball & Ghysels (2017) undersøgte, hvorvidt en MIDAS model kunne opnå en højere forecast præcision end analytikere i forbindelse med estimering af amerikanske selskabers EPS, ved en forecast horisont på et kvartal eller mindre. Ball & Ghysels (2017)

datagrundlag baserede sig på data fra 1984 til 2014, og bestod af 1.050 virksomheder, hvor hovedparten af selskaberne befandt sig i industrierne manufacturing og teknologi. Ball & Ghysels (2017) benyttede hertil en kombination af regnskabstal, herunder varelager, bruttomargin og COGS som inputvariable i modellen, hvilke alle er opgjort på kvartalsbasis. Ydermere benyttede Ball & Ghysels (2017) de pågældende selskabers månedlige aktie merafkast set i forhold til Fama-French's industri porteføljer, aktie volatilitet samt syv makroøkonomiske nøgletal heriblandt oliepriser, inflation og renten på tre måneders T-Bill i USA. Ball & Ghysels (2017) konkluderede på baggrund af resultaterne, at den anvendte MIDAS model opnåede en højere forecast præcision, når spredningen af analytikers estimater var høj, samt når der var tale om mindre selskaber, hvilket dermed bekræfter konklusionerne fremstillet af Brown et al. (1987), samt Bradshaw et al. (2012). Ball and Ghysels (2017) konstaterede yderligere, at en kombination af MIDAS modellen og analytikeres forecast opnåede en bedre forecast præcision end de to forecast separat. Ved anvendelse af analytiker estimater som inputvariable i MIDAS modellen opnåede Ball & Ghysels (2017) en statistisk signifikant bedre performance end analytikerne på ni ud af ti industrier.

Azevedo et al. (2021) videreudviklede studiet af Ball & Ghysels (2017), ved at kombinere to forskellige tilgange til forecast af selskabers indtjening. Azevedo et al. (2021) anvendte Ball & Ghysels (2017) tilgang ved at benytte analytikeres forecast som input i modellen og valgte yderligere at kombinere dette med en regressions model baseret udelukkende på finansiel information. Modellen benyttede blandt andet brutto fortjeneste, hvilket jævnfør Novy-Marx (2013) resultater har en stærk relation til indtjening. Ydermere fulgte Azevedo et al. (2021) Ashton & Wang (2013) tilgang, der viste at ændringer i aktiekurs driver indtjening, ved at inkludere den seneste aktieperformance i modellen. Azevedo et al. (2021) model formår endvidere at outperforme analytikeres forecast præcision samt at opnå et lavere bias end analytikere og andre modeller fremstillet i litteraturen såsom modellen af Hou et al. (2012) samt Li & Mohanram (2014). Babii et al. (2020) valgte ligeledes at videreudvikle Ball and Ghysels (2017) model, ved at introducere en machine learning baseret MIDAS model. Babii et al. (2020) benyttede tilsvarende finansielle og makroøkonomiske inputvariable som Ball and Ghysels (2017), men valgte yderligere at inkludere tekstanalyse af finansielle nyheder med udgangspunkt i Bybee et al. (2020) metode. Babii et al. (2020) resultater indikerede endvidere at implementeringen af en machine learning MIDAS model var bedre til at forecaste price-earnings ratio for amerikanske virksomheder end både analytikere samt random walk modeller, og demonstrerede dermed machine learnings anvendelighed i forbindelse med forecast af selskabers indtjening.

2.2.3 Machine Learning

Anvendelsen af machine learning i den finansielle sektor har de seneste år oplevet en enorm vækst. John Hull understreger dette i nedenstående citat fra Journal of Risk Management in Financial Institutions, hvordan machine learning og data science har vundet indpas på den finansielle sektor:

> More generally, machine learning is used in almost every part of the financial service sector. Not all of my students will become data scientists. But it is becoming increasingly clear that coding skills (ideally Python) and some knowledge of data science are necessary for obtaining and holding down jobs in the finance sector (Hull, 2020, s. 105).

I litteraturen har anvendelsen af machine learning i forbindelse med forecast af selskabers indtjening hovedsageligt fokuseret på indtjeningsændringer, og har dermed forsøgt at bygge videre på forskningen fremstillet af Ou & Penman (1989). Blandt andet undersøgte Baranes & Palas (2019) hvorvidt en support vector machine (SVM) machine learning model var bedre end en logistisk regressionsmodel, på baggrund af 58 finansielle nøgletal. Det fremgik af Baranes & Palas (2019) resultater, at SVM modellen med en gennemsnitlig forecast præcision på 63,4%, var dårligere end en logistisk regression, der opnåede en tilsvarende en forecast præcision på 68,1%. Baranes & Palas (2019) resultater indikerede dog, at SVM modellens forecast præcision var markant bedre end statistiskemodellers i specifikke brancher, herunder særlig i mine og byggebranchen. Hunt et al. (2019) valgte ligeledes at bygge videre på teorien fremsat af Ou & Penman (1989) ved at undersøge hvorvidt machine learning var bedre til at forecaste retningen på næste kvartals indtjeningen. Modsat Baranes & Palas (2019) indikerede Hunt et al. (2019) resultater at en random forest machine learning model formåede at outperforme traditionelle statiske modeller. Hunt et al. (2019) resultater viste endvidere at random forest machine learning modellen opnåede en forecast præcision på 76,8% til sammenligning med 62,3% for en logistisk regressions model. Hunt et al. (2019) konkluderede yderligere, at såfremt price-earnings ratio indgik som inputvariable i modellen ville dette øge forecast præcision fra 76,8% til 79,8%, hvilket endvidere bekræfter Ashton & Wang (2013) argument om, at ændringer i aktiekurs er forklarende for selskabers indtjening.

Xinyue et al. (2021) replikerede både Hunt et al. (2019) og Ou and Penman (1989) metode, men benyttede derimod en LighGBM machine learning model. Endvidere viste LighGBM modellen en forecast præcision på 64,2%, og outperformede dermed Ou and Penman (1989) model. LighGBM modellen formåede dog ikke at outperforme Hunt et al. (2019) random forest model med en forecast præcision på 76,8%. Den relativt begrænsede litteratur inden for forecast af indtjeningsændringer indikerer, at kun visse machine learning modeller har formået at outperformet statistiske modeller.

Andre studier har beskæftiget sig med anvendeligheden af machine learning, i forbindelse med forecast af selskabers fremtidige indtjening i en regressionssammenhæng frem for i en klassificeringssammenhæng, hvilket hidtil er beskrevet. Van Binsbergen et al. (2020) konstruerede blandt andet en machine learning model til forecast af selskabers fremtidige indtjening, ved at benytte historiske regnskabstal, makroøkonomiske variable samt analytikeres konsensus estimat som inputvariable. Van Binsbergen et al. (2020) undersøgte forecast præcisionen på fem forskellige forecast horisonter, herunder på et kvartal, to kvartaler, tre kvartaler, et år og to års forecast horisont. Resultaterne heraf viste endvidere, at den anvendte machine learning model opnåede en signifikant bedre forecast præcision end analytikere på alle tidshorisonter. Ydermere kunne Van Binsbergen et al. (2020) konstatere, at analytikeres forecast var overoptimistiske på alle forecast horisonter, hvilket understøtter tidligere fremstillede litteratur omkring optimisme bias jævnfør afsnit 2.1.2. Van Binsbergen et al. (2020) resultater indikerede ydermere, at analytikeres forecast præcision forringes ved forlængelse af forecast horisonten og er dermed konsistente med konklusionerne fremstillet af Brown, Richardson, et al. (1987) og Bradshaw et al. (2012). Særlig bemærkelsesværdigt ved Van Binsbergen et al. (2020) model er, at analytikeres estimater udgjorde den inputvariabel med størst forklaringskraft. På baggrund af studiet er det dermed ikke muligt endegyldigt at konkludere, hvorvidt machine learning model performer bedre end analytikere.

Litteraturen inden for machine learning har de seneste år skiftet fokus fra de mere traditionelle machine learning modeller, til mere komplekse deep learning modeller. I litteraturen er deep learning modeller anvendt til identificering af regnskabssvindel (Alghofaili et al., 2020; Craja et al.,

2020), valutakurs forecast (Dixon et al., 2017; Sermpinis et al., 2012), aktiekurs forecast (Shen & Shafiq, 2020; Vargas et al., 2017) samt i forbindelse med adskillelige andre problemstillinger. Sezer et al. (2020) kunne i et omfattende litteratur review af deep learnings anvendelse inden for finansielle tidserier konkludere, at der i litteraturen hovedsageligt benyttes Recurrent Neural Networks (RNN), Deep Multi Layer Perceptron (DMLP) og Convolutional neural network (CNN). Ydermere konstaterede Sezer et al. (2020) af LSTM var den mest populære RNN-model, hvilket skyldes modellens evne til at håndtere tidserier effektivt samt den gode performance sammenlignet med andre deep learning modeller. Elend et al. (2020) er det eneste identificerede studie i litteraturen, der benytter deep learning i forbindelse med forecast af selskabers fremtidige indtjening. Elend et al. (2020) benyttede regnskabstal, såsom aktiver, omsætning og vareforbrug som inputvariable i sine to deep learning modeller, henholdsvis LSTM og TCN. Ydermere indgik aktiekurser og volume de seneste 20 dage som inputvariable i modellen, med henblik på at inkorporere markedets forventninger til aktien og dermed virksomhedens evne til at generere indtjening. Inkludering af markedsbaserede indikatorer, har tidligere vist sig at forbedre performance af machine learning modeller i forbindelse med indtjeningsforecast. Elend et al. (2020) evaluerede den endelige performance på baggrund af seks måneders test data i år 2017, hvortil TCN og LSTM formåede at outperforme analytikere med henholdsvis 13,2% og 12,2% med en statistisk signifkans, opgjort ved performancemålet MSE. Elend et al. (2020) resultater er revolutionerende, da studiet skaber evidens for at machine learning udgør en bedre proxy for selskabers fremtidige indtjening end analytikere, der hidtil blev ansat som uovertrufne på korte forecast horisonter.

3 Metode

I følgende kapitel uddybes det metodiske grundlag for afhandlingen. Kapitlet opdeles i henholdsvis en beskrivelse af data grundlag, datakilder samt data cleaning. Beskrivelsen af datagrundlaget omhandler tidshorisont samt forudsætning for datagrundlaget. I afsnittet omhandlende datakilder, foretages en beskrivelse af kilder samt overvejelser bag valget af disse kilder.Afslutningsvis præsenteres data cleaning processen, hvori datagrundlaget renses således, at det kan benyttes som input i en machine learning model.

Key takeaways: Det primære datagrundlag, der benyttes som input i machine learning modellen er hentet fra IBES databasen, CRSP databasen, Compustat samt Fama & French hjemmeside. En omfattende data cleaning af datagrundlaget inkluderede imputing, scalling samt behandling af outliers, og resulterede i et endeligt datasæt bestående af 128.668 observationer fra 3.754 virksomheder i alt.

3.1 Data

Følgende afsnit har til formål at give et dybdegående indblik af afhandlingens datagrundlag og dertilhørende forudsætninger, samt at skabe transparens i forhold til metodiske valg foretaget i forbindelse med databehandling, med henblik på at sikre reproducerbarheden af resultater.

Liu et al. (2020) fremhæver markante problemstillinger relateret til akademisk forskning inden for machine learning, herunder vigtigheden af replikerbarhed og reproducerbarhed, hvilket forskere ofte ignorerer ifølge Liu et al. (2020). Udviklingen af machine learning algoritmer og metodiske tilgange relateret til anvendelsen af Machine Learning er under konstant forandring. Af denne årsag er det fair at antage at metodiske valg foretaget i tidligere forskning muligvis ændres såfremt forskningen foretages i dag, hvilket kun er muligt såfremt forskningsresultater fra tidligere studier er replikerbare og reproducerbare. Ifølge Liu et al. (2020) trues validiteten af Machine Learning modeller anvendt i forskning såfremt vigtigheden af replikerbarhed og reproducerbarhed ignoreres.

3.1.1 Beskrivelse af datagrundlaget

Datagrundlaget i afhandlingen baserer sig på fra data fra IBES, COMPUSTAT, CRSP og Fama French. For at undgå time period bias består datagrundlaget af data fra perioden 1995-2020 og udgøres af 552.258 antal observationer fordelt på 15.926 selskaber[4] før data cleaning (CFI Institute, 2021). Som belyst i afsnit 1.1.1 danner datagrundlaget et repræsentativt grundlag for analysen samt generaliserbarhed i konklusioner.

[4]Udledt på baggrund af COMPUTATS unikke virksomheds identifikator (GVKEY)

I tilstræbelsen på at sikre et validt sammenligningsgrundlag mellem analytikere og Machine Learning modelens forecast præcision, er det essentielt sikre at den endelige machine learning model der anvendes i afhandlingen, ikke besidder information der ikke var tilgængeligt for analytikere på det givne tidspunkt og omvendt, også betegnet look-ahead bias[5]. For at undgå look-ahead bias i afhandlingen opstilles følgende forudsætninger for datagrundlaget:

1) Machine Learning modeller vil have en forecast horisont på henholdsvis 1, 2 og 4 kvartaler og vil ikke anvende information, der forekommer efter forecast tidspunktet.

2) Information relateret til regnskabstal og EPS, der anvendes som input i Machine Learning modellerne baserer sig på data et kvartal forud for forecast tidspunktet (q_{t-1}).
Årsagen til dette er, at kvartalsregnskaber først offentliggøres op til 90 dage efter endt kvartal, hvorfor regnskabsinformation relateret til indeværende kvartal (q_t) ikke er tilgængeligt på forecast tidspunktet (U.S securities and exchange commision, 2021).

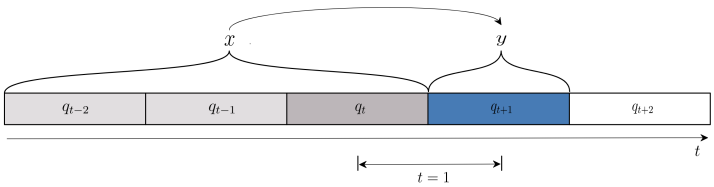
3) Analytikerestimater hentes fra IBES databasen på samme dato som Machine Learning modellernes forecast tidspunkt. Dette sikrer, at hverken Machine Learning modellerne eller analytikerne besidder en informations fordel på forecast tidspunktet.

4) Aktierelateret data udgøres af realtidsdata, og anvendes derfor frem til datoen for afgivelsesestimatet. Aktiedata vil endvidere benyttes for de seneste 30 dage op ledende op til forecast tidspunktet.

5) Forudsætning omkring aflæggelse af regnskab ultimo i månederne; marts, juni, september og december.

For at illustrere de fremstillede forudsætninger samt dertilhørende indvirkning på datagrundlaget, tages der udgangspunkt i et eksempel fra datasættet, ved en forecast horisont på 1 kvartal.

Figur 3.1: Forecast af EPS med machine learning - tidslinje



Note: Figuren illustrerer machine learning forecast af selskabers fremtidige EPS på baggrund af kvartalsregnskaber, q_t, på tidspunkt, t. Ved brug af historisk data, x, estimeres fremtidig EPS, y.

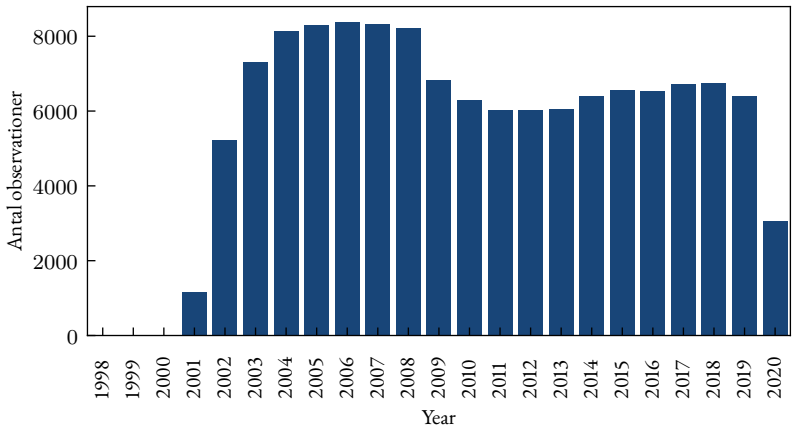
I figur 3.1 tages udgangspunkt i selskabet Apple, hvortil forecast tidspunktet udgør slutningen af 2. kvartal 2019 (q_t) og vi hertil ønsker at estimere Apples EPS 3. kvartal 2019 (q_{t+1}). Machine learning modellens inputvariable udgøres hertil af regnskabsmæssige nøgletal og aktiedata. Apples kvartalsregnskab for kvartal 2 (q_t) blev først offentliggjort den 30. juli 2019 og er dermed ikke tilgængeligt på forecast tidspunktet (q_t). Jævnfør forudsætning 2) baseres Apples regnskabsmæssige information derfor på data kvartalet forud for forecast tidspunktet, det vil sige 1. kvartal 2019 (q_{t-1}). Modsat regnskabsinformation anvendes aktierelateret data fra den pågældende måned for forecast tidspunktet afgives jævnfør forudsætning 4), hvilket i eksemplet betyder at aktiedata for juni 2019

[5]Look-ahead bias forekommer, når information der ikke var tilgængeligt eller kendt i analyseperioden benyttes.

benyttes. For at sikre at den machine learning model, der appliceres, ikke har informationsmæssig fordel og omvendt benyttes analytiker estimater fra samme tidspunkt som forecast tidspunktet jævnfør forudsætning 3), det vil sige fra slutningen af 2. kvartal 2019 (q_t) svarende til den 30. juni 2019.

I forbindelse med træning af machine learning modeller samt evaluering af forecast præcision i forhold til analytikere er det essentielt, at der foreligger et reelt EPS for det givne kvartal. På baggrund heraf er observationer, hvorved der ikke foreligger etkvartalsmæssigt EPS fjernet fra datagrundlaget. I figur 3.2 fremgår distributionen af observationer i datagrundlaget fordelt på de respektive kvartaler i perioden 1995-2020. Figuren baserer sig endvidere på datagrundlaget efter foretaget data cleaning. En detaljeret gennemgang af data cleaning processen samt dertilhørende overvejelser fremgår af efterfølgende afsnit 3.2.

Figur 3.2: Antal observationer i datagrundlaget



Note: Figuren illustrerer antallet af observationer i datagrundlaget fordelt på hvert år i perioden 1995-2020.

Jævnfør figur 3.2 består datasættet af 128.668 antal observationer i perioden 1998-2020, til trods for, at dataudtræk fra de primære datakilder baserer sig på perioden 1995-2020. I den resterende del af perioden er observationerne omtrent ligeligt fordelt på de respektive år, med et gennemsnit på cirka 7.000 observationer per år. Til trods for de manglende observationer i perioden 1995-1999 vurderes datagrundlaget at danne et validt grundlag for resultater samt generaliserbarhed i konklusioner, da datagrundlaget strækker sig over en tilstrækkelig lang periode til at tage højde for ændringer i underliggende økonomiske konjunkturer.

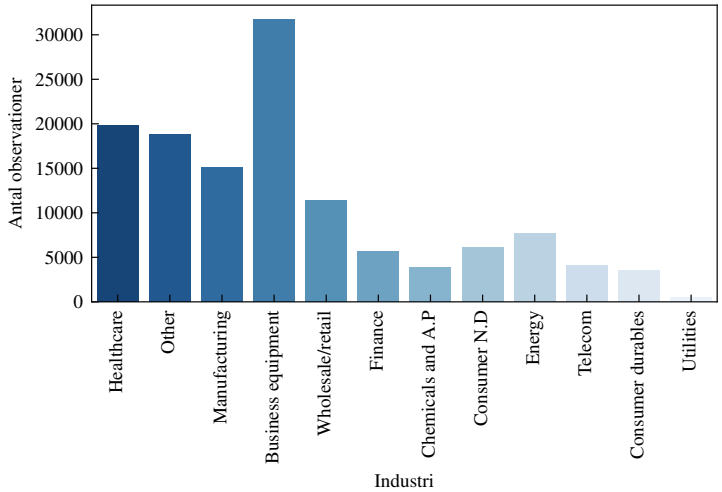
I forbindelse med besvarelsen af arbejdsspørgsmål 2, der omhandler hvorvidt, der er forskel i forecast præcision mellem analytikere og en machine learning model på tværs af industrier, er det essentielt at sikre et tilstrækkeligt datagrundlag for hver enkelt industri. Endvidere er hver enkelt observation i datagrundlaget inddelt i en industri med udgangspunkt i de respektive SIC koder, på baggrund af 12 industrier defineret af Fama & French (2021). Det er endvidere fundet hensigtsmæssigt at inddele i 12 industrier, da denne inddeling muliggør analyse af finansielle virksomheder isoleret, hvilket tidligere studier jævnfør afsnit 2.2.2 enten udelukkede af analysen eller identificerede en markant dårligere forecast præcision sammenlignet med andre industrier. Ydermere sikrer inddelingen i 12 industrier et tilstrækkeligt antal observationer for industrierne, hvilket er essentielt i forbindelse med træning af machine learning modeller. En yderligere fragmentering af industrier såsom Fama & French 48 industriinddeling, ville resultere i færre observationer for hver enkelt industri, hvilket alt

andet lige vil forringe forecast præcisionen i de anvendte machine learning modeller grundet mindre træningsdata per industri.

Ifølge (Bhojraj et al., 2003) sikrer industriinddelingen på baggrund af Fama & French (2021) en mere retvisende inddeling i industrier sammenlignet med en inddeling foretaget udelukkende på baggrund af SIC koder. Ifølge Bhoraj et al (2003) er det ligeledes muligt at benytte GICS og NAICS koder til inddeling af selskaber i specifikke industrier. Endvidere er det grundet en mangel på industriafkast opgjort på baggrund af GICS og NAICS, fundet hensigtsmæssigt at anvende Fama & French (2021) industriopdeling.

I figur 3.3 fremgår distributionen af observationer fordelt på industrier efter foretaget data cleaning jævnfør efterfølgende afsnit 3.2. Ydermere foreligger en detaljeret beskrivelse af de 12 respektive industrier i bilag B. På baggrund af figuren, kan der konstateres en markant forskel i antallet af industrier repræsenteret i datagrundlaget. *Business Equipment* industrien udgør den største industri med cirka 25% af observationerne mens *Utilities* industrien udgør den mindste industri med kun 0,3% af de samlede observationer i datagrundlaget. Forfatterne er opmærksomme på, at andelen af observationen i de respektive industrier kan have indflydelse på machine learning modellers forecast præcision, og vil hertil løbende i analysen vurdere hvorvidt visse industrier bør udelades.

Figur 3.3: Observationer fordelt på industrier



Note: Figuren illustrerer antallet af observationer fordelt på industrier.

3.1.2 Datakilder

I følgende afsnit vil de primære kilder benyttet til at skabe afhandlingens datagrundlag præsenteres samt evalueres i forhold til kildernes reliabilitet. I afhandlingen benyttes fire datakilder for at opnå de forskellige typer af data, der er nødvendige for analysen: 1) Institutional Brokers Estimate System (IBES) er benyttet for at fremskaffe et standardiseret og justeret earnings per share, 2) Compustat/CRSP Merged blev benyttet for at opnå selskabers fundamentale regnskabstal, 3) CRSP blev anvendt til aktiedata, herunder månedligt aktieafkast, udestående aktier og volume (CRSP, 2021d)og 4) Fama & French Industriafkast blev benyttet til at udregne merafkastet i forhold til industrien. En udtømmende liste over variable hentet fra de fire ovenstående kilder fremgår af bilag A.1.

Ibes

Institutional Brokers Estimate System (IBES) blev benyttet til at fremskaffe analytikeres EPS estimater, samt realiseret EPS, og blev tilgået via Refinitivs Eikon API løsning (IBES & Refinitiv, 2021). I programmeringssproget Python blev et script udviklet jævnfør bilag D, der ved anvendelse af Eikons API gjorde det muligt hente det ønskede data. IBES databasen ejes og drives i dag af Refinitiv, og indeholder informationer på mere end 22.000 selskaber i 90 forskellige lande samt data fra mere end 18.000 analytikere (Refinitiv, u.å.). IBES databasen er den mest anvendte samt omfattende database inden for akademia. I afhandlingen benyttes data fra både IBES actual databasen, samt IBES summary databasen. Hertil benyttes både estimeret EPS og realiseret EPS i standardiseret form, hvor Refinitiv har korrigeret for at EPS bliver opgjort på samme grundlag. Dette er med til at sikre konsistens i data, da nogle analytikere eksempelvis udelader ekstraordinære omkostninger, mens andre inkluderer disse. Forfatterne har valgt at benytte realiseret EPS fra IBES databasen fremfor GAAP EPS fra Compustat for at sikre et konsistent sammenligningsgrundlag mellem realiseret EPS og analytiker estimater.

Til trods for, at IBES databasen er den mest benyttede database inden for analytiker estimater, samt den primære kilde til estimater inden for akademia, er adskillelige problematikker forbundet med databasen identificeret over tid. Payne & Thomas (2003) identificerede hertil et problem i, at IBES databasen kun rapporterer justerede EPS værdier korrigeret for aktiesplit med to decimaler, hvilket betyder at nogle analytiker estimater fremstår med en estimatfejl på 0, hvor den reelle estimatfejl var ikke 0. Payne & Thomas (2003) anbefaler hertil at research, der særligt fokuserer på nulfejl estimater og karakteristika, heraf bør være særligt opmærksomme på denne fejl. I afhandlingen er endvidere det ikke fundet nødvendigt at tage højde for denne problematik, da analytikerestimaterne anvendt i afhandlingen har en forecast horisont på henholdsvis et, to og fire kvartaler ude i fremtiden, hvorfor sandsynligheden for forecast fejl på 0 er minimal. Ydermere er adgangen til IBES data gennem Eikon forholdsvis begrænset, hvorfor det ikke er muligt at hente de ikke justerede analytikerestimater. Et andet potentielt problem med IBES databasen er ændringen i data med tilbagevirkende kraft, hvilket dermed vil kunne påvirke reproducerbarheden af vores analyse. Ljungqvist et al. (2009) fandt frem til at IBES data downloadet i perioden 2000 til 2007 matchet på observationer adskilte sig fra hinanden med mellem 1,6% (6.580 observationer) og 21,7% (97.582 observationer), hvor ændringerne inkluderede ændring i aktie anbefalinger, nye estimater og fjernelse af tidligere estimater. Call et al. (2016) fandt frem til, at ændringen primært fandt sted i IBES detailed og ikke

IBES summary, hvilket benyttes i afhandlingen, hvorfor denne problematik ikke vil udgøre et problem for reproducerbarheden og kvaliteten af afhandlingens datagrundlag.

Compustat North America

Ou & Penman (1989) samt andre studier, der baserede sig på deres metode, fokuserede på ratios, hvorimod datagrundlaget i afhandlingen, vil benytte regnskabstal. Valget af specifikke regnskabstal i afhandlingen er primært baseret på regnskabstal inkluderet i Elend et al. (2020), og yderligere på de mest signifikante regnskabstal fra Ball and Ghysels (2017). En udtømmende liste af regnskabsvariable anvendt i afhandlingen fremgår af bilag A.1. Regnskabsdata benyttet i afhandlingen er hentet fra Computstat, hvilket er konsistent med litteraturen, hvor både Ou and Penman (1989) og Elendt et al. (2020) benytter Compustat. Regnskabsdata er hentet via CRSP/Compustat Merged gennem WRDS, da dette endvidere muliggør en kombination af regnskabstal fra Compustat med aktiedata fra CRSP på baggrund af PERMNO[6] (CRSP & S&P Capital IQ, 2021).

Compustat er en fundamental database udviklet og vedligeholdt af S&P Capital IQ, der indeholder information på mere end 24.000 aktive og inaktive børsnoterede virksomheder, med kvartals data dateret tilbage til 1962 (Sonne, 2021). Compustat er anerkendt for den høje datakvalitet, hvor både WRDS og akademikere generelt finder meget få fejl i databasen (WRDS, 2021).

Crsp

Adskillelige studier (W. Beaver et al., 1980; W. H. Beaver et al., 1987; Collins et al., 1987; Freeman, 1987) har identificeret hvordan aktieafkast samt aktiekurser opfanger information omkring fremtidig indtjening før tid. Derudover har studiet af eksempelvis Ball and Ghysels (2017), Elend et al. (2020) samt Hunt et al (2019) vist at inkludering af aktiedata såsom afkast, handelsvolumen mfl. bidrog positivt til at forudsige både retningen og niveauet af den fremtidige EPS. I afhandlingen vil aktiedata ligeledes inkluderes i datagrundlaget, hvilket endvidere er hentet gennem CRSP databasen. CRSP er den mest anerkendte og hyppigst anvendte aktiedatabase inden for forskning. Mere end 500 akademiske institutioner fordelt over 35 lande, benytter CRSP i forbindelse med forskningsprojekter (CRSP, 2021c). CRSP blev startet i 1963 af University of Chicago Booth School of Business, og databasen dækker i dag mere end 32.000 aktive og inaktive virksomheder med dagligt data tilbage til 1925 på NYSE, NASDAQ og Bats (CRSP, 2021a). Adgangen til CRSP foregår gennem WRDS, hvor aktiedata de seneste 30 dage før forecast tidspunktet anvendes, som beskrevet i afsnit 6.2.1 (CRSP, 2021d).

Fama French Industri afkast

På baggrund af klassificeringen med udgangspunkt i Fama & French (2021) 12 industrier, som belyst i afsnit 6.2.1, er det fundet hensigtsmæssigt at inkludere det specifikke industriafkast. Industriklassificeringen muliggør udregningen af et merafkast mellem de respektive virksomheder og den pågældende industri de befinder sig i. Industriafkastet er hentet på Fama & French hjemmeside, hvortil beregninger baserer sig på CRSP-databasen (Fama & French, 2021). Merafkastet i forhold til

[6] PERMNO er en unik identifikator på aktieniveau.

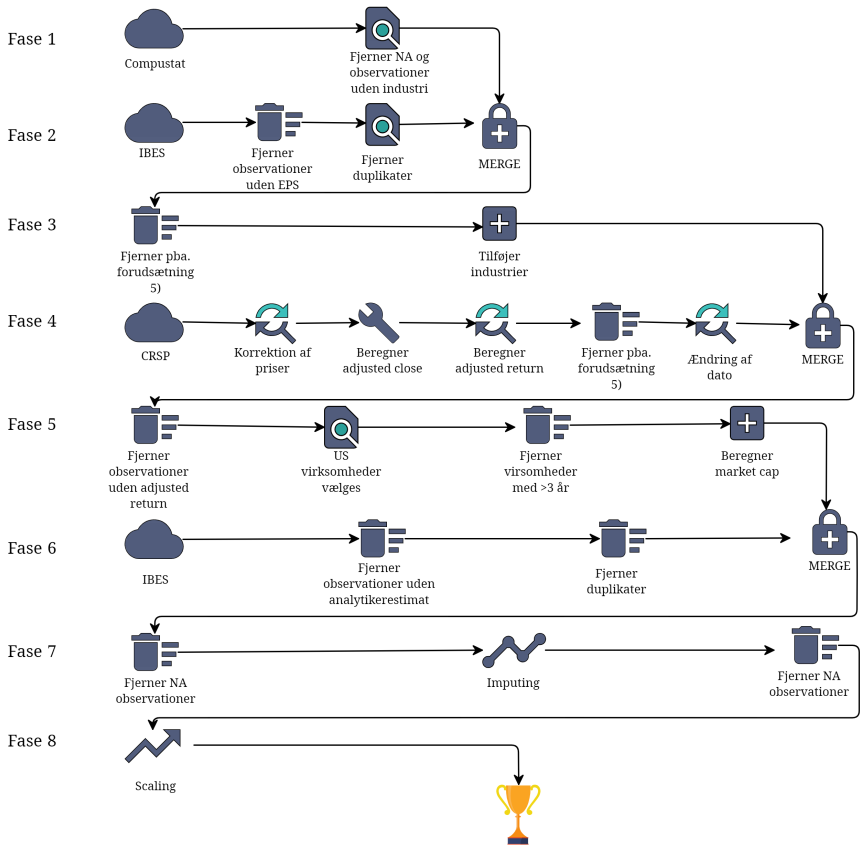
industrien benyttes ligeledes af Ball & Ghysels (2017) som inputvariabel i MIDAS modellen. Argumenterne for at inkludere et merafkast i forhold til industrien er, at dette giver mulighed for at isolere den enkelte virksomheds performance, renset for en industrieffekt. Dette vil alt andet lige give en bedre indikation selskabernes fremtidige indtjening, da et højt aktieafkastet alene i høj grad kan været præget af, en øget efterspørgsel på aktier i den givne industri en given periode.

3.2 Data cleaning

Data cleaning udgør et grundlæggende element i data science og er en essentiel proces, der foretages forud for applicering af machine learning modeller. Data cleaning henviser til processen hvori ukorrekte, ufuldkommen samt irrelevant information og manglende data identificeres og derefter modificeres, erstattes eller slettes baseret på væsentlighedsgraden (Tableau Software LLC., 2021). Mangel på data cleaning kan forvrænge resultater og i værste fald ændre udfaldet af konklusioner baseret på det underliggende datagrundlag (Chu et al., 2016). I følgende afsnit vil en dybdegående gennemgang af data cleaning samt metodiske overvejelser i forbindelse med datagrundlaget gennemgås.

Grundet omfanget af datagrundlaget målt på henholdsvis antallet af observationer samt kolonner data, er det fundet hensigtsmæssigt at foretage data cleaning i programmeringssproget Python. Data cleaning processen foretaget i afhandlingen er fremgår af figur 3.4, der illustrerer de mest afgørende elementer af data cleaning processen i kronologisk rækkefølge. Data cleaning processen er endvidere inddelt i forskellige faser med henblik på at gøre det mere overskueligt for læseren at forstå de enkelte delelementer i processen. For at undgå gentagelser i beskrivelsen af data cleaning processen på tværs af forecast horisonterne er processen beskrevet med udgangspunkt i en forecast horisont på 1 kvartal, da processen er identisk på tværs af alle forecast horisonter.

Figur 3.4: Data cleaning proces



Note: Figuren illustrerer de forskellige faser i data cleaning processen.

Datagrundlaget består i sin grundform af fem forskellige datainputs fra de fire forskellige datakilder; Compustat, CRSP, IBES of Fama & French jævnfør figur 3.4. For at muliggøre anvendelsen af machine learning modeller er det hertil essentielt, at de fire separate datainput kombineres til et samlet datasæt, bestående af de relevante informationer fra de respektive datakilder.

3.2.1 Fase 1

I fase 1 blev fundamentale regnskabstal hentet fra Compustats database. Det fremgik hertil, at hovedparten af data vedrørende regnskabsposterne *Restructuring cost* og *SG&A* ikke var tilgængelig, hvorfor disse inputvariable blev fjernet fra datasættet. Envidere fremgik det, at en del af observationerne i datasættet manglede en *SIC*[7] kode. Da vi jævnfør arbejdsspørgsmål 2 ønsker at analysere forskelle i en machine learning models forcast præcision på tværs af industrier, blev observationer med manglende SIC koder ligeledes fjernet fra datasættet.

[7] Standard Industrial Classification (SIC) er firecifrede koder, der kategoriserer de industrier som virksomheder tilhører (Kenton, 2019).

3.2.2 Fase 2

På baggrund af det korrigerede data fra Compustat, var det muligt at generere en liste over de selskaber, der indgik i datasættet i form af unikke *CUSIP*[8] numre. Listen med CUSIP-numre blev hertil anvendt som selekteringskriterie i dataudtrækket fra Eikon, indeholdende de historiske realiserede EPS værdier. Ved gennemgang af dataudtrækket fremgik det, at en større andel af observationerne havde manglende værdier i realiseret EPS kolonnen. Da det, som belyst i afsnit 6.2.1, er afgørende i forbindelse med træning af machine learning modeller samt validering af resultater, at der for hver enkelt observation foreligger et realiseret EPS, er observationer med manglende realiseret EPS værdier derfor ekskluderet fra datasættet.

Ved inspektion af de respektive observationers realiserede EPS værdier blev der identificeret fejl i datasættet i form af duplikater. For visse observationer fremgik identiske realiseret EPS værdier sekventielt for de givne selskaber. Til trods for, at det er muligt for selskaber at have identisk EPS på forskellige tidspunkter, kunne det konstateres, at der i tilfældene var tale om systematiske fejl. Endvidere havde observationer, hvor de identiske realiseret EPS værdier forekom gentagne gange for samme selskab, forskellige datoer for *Quarter end date* og *Calculation date*, der henviser til henholdsvis det kvartal realiseret EPS vedrører samt datoen for beregning. Denne problematik er ikke hidtil beskrevet i litteraturen, men indikerer at Eikon, såfremt realiseret EPS ikke er tilgængeligt, opgiver den sidst tilgængelige realiseret EPS værdi. Da duplikater af realiseret EPS ikke afspejler virkeligheden og potentielt påvirker afhandlingens resultater er disse observationer ekskluderet fra datasættet. Efter at have korrigeret for duplikater i datasættet merges realiseret EPS datasættet med regnskabsdata fra Compustat fra fase 1, på baggrund af et inner join på henholdsvis *CUSIP* og *Quarter End Date*.

3.2.3 Fase 3

I datasættet er der hidtil ikke korrigeret for, at selskaberne skal aflægge kvartalsregnskab i månederne; marts, juni, september og december, jævnfør forudsætning 5) i afsnit 6.2.1. Efter at have kombineret regnskabsdata og realiseret EPS data i fase 2 blev observationer hvor kvartalsregnskabet ikke aflægges i de specificerede måneder derfor fjernet. Endvidere er hver enkelt observation i datasættet klassificeret i en af de 12 industrier defineret af Fama & French's industriklassifikation på baggrund af de respektive *SIC* koder jævnfør bilag B.

3.2.4 Fase 4

I denne fase blev månedlige aktiedata hentet fra CRSP databasen på baggrund af listen med *CUSIP*-numre fra fase 1. Da det jævnfør afsnit 3.1.2 ønskes at beregne et månedligt aktieafkast for at muliggøre sammenligning med industriafkastet for de pågældende industrier, var det hertil nødvendigt at beregne en *adjusted price*, for at korrigere aktiekursen opgivet af CRSP for aktiesplit og udbytte.

[8] CUSIP numre er ni-cifret numerisk eller ni-tegns alfanumerisk kode, der identificerer et nordamerikansk finansiel instrument med henblik på at lette identifikation i clearing og handel.

Endvidere var det før beregning nødvendigt at korrigere de oplyste aktiekurser for negative værdier. Såfremt en aktiekurs ikke er tilgængelig en given dato benytter CRSP gennemsnittet af den pågældende akties bid/ask kurs, og oplyser dette som en negativ værdi (CRSP, 2021b).For at korrigere de negative værdier blev alle aktiekurser oplyst i datasættet derfor ændret til absolutte værdier. Den nye variabel *Adjusted price* blev beregnet ved at dividere *Price Alternate* med *Cummulative Adjustment Factor* tilhørende den enkelte observation. *Price Alternate* benyttes fremfor *Price*, da denne afspejler den senest tilgængelige aktiekurs og dermed giver mulig for at inkludere observationer, hvor aktiekursen af en given årsag ikke var tilgængelig den sidste dag i måned (CRSP, 2021b). Endvidere var det på baggrund *adjusted price* muligt at beregne et månedligt afkast for hver observation i datasættet.

Grundet forudsætning 5) jævnfør afsnit 6.2.1, fjernes observationer fra datasættet, hvor kvartalsregnskabet ikke aflægges i de specificerede måneder. Efter at have korrigereret datasættet fra CRSP blev datasættet merget med data fra fase 3, på baggrund af et inner join på henholdsvis *PERMNO* og *Calculation date*.

3.2.5 Fase 5

Efter at have merget data fra de forskellige datainputs i de tidligere beskrevne faser, fremgik det endvidere, at visse observationer havde manglende værdier i det beregnede månedlige afkast. Da det jævnfør afsnit 3.1.2 ønskes at sammenligne aktieafkastet for de enkelte observationer med et industriafkast for de pågældende industrier, blev observationer hvor det ikke var muligt at beregne et månedligt aktieafkast derfor fjernet fra datasættet.Hidtil havde observationerne i det kombinerede datasæt, som belyst i afsnit 1.1.1, baseret sig på selskaber, der er børsnoterede i USA, hvorfor datasættet både indeholder amerikanske og ikke-amerikanske selskaber. Et eksempel herpå er Novo Nordisk, der indgår i datasættet da selskabet er børsnoteret både på NASDAQ i Danmark samt på New York Stock Exchange. Da vi i afhandlingen udelukkende ønsker at undersøge amerikanske selskaber jævnfør afsnit 1.1.1, blev ikke-amerikanske selskaber derfor ekskluderet fra datasættet. Ekskluderingen blev foretaget på baggrund af kolonnen *HQ country.*

I tilstræbelsen på at sikre, at en machine learning model kan identificere mønstre og tendenser i det underliggende datagrundlag, er det essentielt, at der foreligger en tilstrækkelig mængde observationer af hver virksomhed i datasættet. Af denne årsag er virksomheder hvortil færre end tre års data forekommer fjernet fra datasættet.

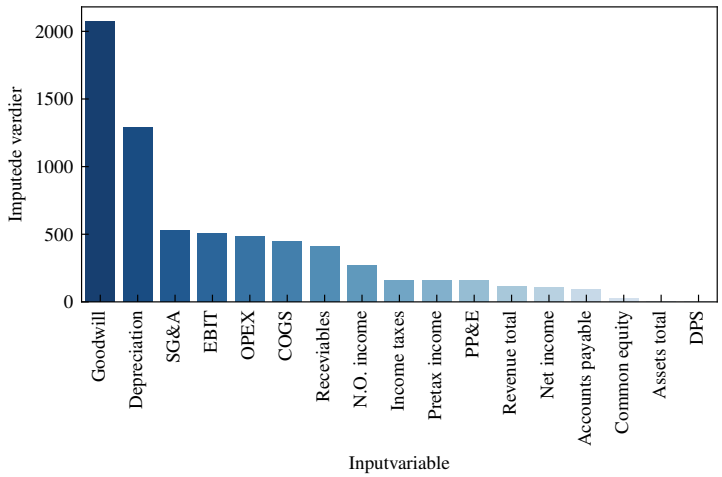
3.2.6 Fase 6

I denne fase blev de sidste informationer inkluderet i datasættet, i form af analytikeres EPS estimater. Endvidere blev analytikerestimaterne hentet fra IBES databasen på baggrund af *CUSIP* listen genereret i fase 1, og blev hertil merget med de forskellige datainputs fra foregående fase, ved et left merge på henholdsvis *CUSIP* og *Calculation date.* Alle datainputs blev hermed kombineret til et samlet datasæt, hvilket var et af formålene med data cleaningen processen. Ved gennemgang af det samlede datasæt fremgik det, at visse observationer manglede værdier for alle regnskabstal. Grundet omfanget af manglende værdier blev det fundet mest hensigtsmæssigt at slette disse værdier fra det samlede datasæt.

3.2.7 Fase 7 (Imputing)

På baggrund af data cleaning proces eksisterede der stadig observationer med adskillelige manglende værdier på primært *SG&A*, *Goodwill* og *Depreciations*. Manglende værdier udgør et væsentligt problem, da machine learning modeller ikke fungerer, såfremt data indeholder manglende værdier. I litteraturen er adskillige fremgangsmetoder fremstillet, der beskriver håndteringen af manglende værdier, lige fra at erstatte manglende med 0 til et væld af simple såvel som avancerede imputation metoder. Imputation henviser til, at der gennem statiske og matematiske modeller forsøges at give det bedste estimat på den reelle værdi (Drechsler, 2020). I litteraturen er der identificeret adskillige metoder til imputing af tidsserie data. Honaker & King (2010) udvikler i deres paper en multipel imputation model baseret på en bootstrappet Expection Maximization algorithme, der er implementeret i pakken Amelia II i programmeringssproget R. Amelia II har opnået lovende resultater med en høj robusthed på multivariat tidserie data (Bauer et al., 2017). Andre eksempler på imputation metoder anvendt på tidserie data er PCA og beslægtede algoritmer (John et al., 2019), Deep Learning (Fang & Wang, 2020) samt diverse Expectation Maximization metoder (Arcadinho & Mateus, 2019; Dong & Peng, 2013). Førnævnte Amelia II imputation metode er testet på afhandlingens datagrundlag, hvortil det kunne konstateres at kompleksiteten af datasættet medførte at Amelia II fejlede, da det ikke kunne tage højde for trenden for hver virksomhed. Det blev i stedet fundet hensigtsmæssigt at anvende en mindre matematisk kompleks model i form af lineær interpolation, der grundlæggende forsøger at forbinde to pointer i datasættet med hinanden. Den manglende værdi interpoleres hertil med en værdi på denne linje (Bayen & Siauw, 2015). Implementeringen af lineær interpolation blev foretaget ved brug af Python biblioteket Pandas *Interpolate* funktion. I funktionen blev interpoleringen begrænset til maksimalt at interpolere 2 sekventielle værdier, samt at manglende værdier skulle være omgivet af komplet data (Reback et al., 2021). Begrænsninger er valgt for at sikre en så nøjagtig interpolering som muligt samt for at mindske graden af information bias. I figur 3.5 illustreres interpolerede antal datapunkter fordelt på de respektive inputvariable i datasættet. Det fremgår hertil, at 6.835 datapunkter er interpoleret, hvortil størstedelen af interpoleringerne er foretaget for inputvariablene *Depreciation* og *Goodwill*.

Figur 3.5: Imputede inputvariable



Note: Figuren illustrerer antallet af imputede værdier fordelt på de forskellige inputvariable.

3.2.8 Fase 8 (Scaling)

En væsentlig udfordring ved anvendelsen af tværsnit tidserie data i machine learning er skalaforskelle. Dette kommer blandt andet til udtryk i datasættet i form af mindre selskaber med regnskabstal i opgjort i millioner og større selskaber, hvor disse er opgjort i milliarder. For at skabe et validt sammenligningsgrundlag på tværs af selskaber, som en machine learning model kan forstå, er det nødvendigt at skalere data.

Yan & Zheng (2017) valgte i forbindelse med estimering af fremtidigt aktieafkast at skalere data på baggrund af 15 forskellige variable, herunder aktiver, omsætning, markedsværdi mfl. Det viste sig endvidere at den statistiske signifikans af Yan & Zheng (2017) resultater varierede væsentligt fra de forskellige kombinationer af regnskabstal og skaleringsvariable, hvorfor det ikke var muligt at konkludere hvilket skaleringsvariabel der var mest optimal. I afhandlingen er det fundet hensigtsmæssigt at skalere datagrundlaget på baggrund af værdien af totale aktiver, hvilket er identisk med fremgangsmetoden i studiet af Elend et al. (2020). Skaleringen af datagrundlaget er baseret på formel 3.1.

$$X_{skaleret} = \frac{X}{\max(1, \text{Totale aktiver})} \tag{3.1}$$

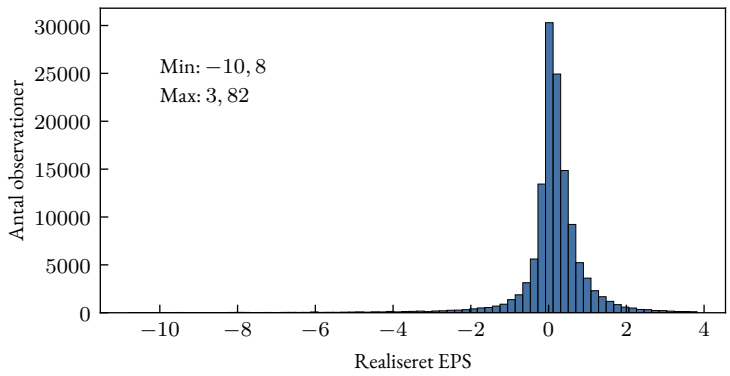
3.2.9 Behandling af outliers

Den omfattende data cleaning foretaget i afsnit 3.2 har sikret korrektion af væsentlige fejl og mangler i data fra de fire datainputs, og giver dermed et solidt fundament for afhandlingens analyse. Endvidere har data cleaning processen sikret, at der ikke forekommer manglende værdier, hvilket er et krav i forbindelse med implementering af en machine learning model i kapitel 5.

Forud for anvendelsen af datagrundlaget i en machine learning model er det essentielt at tage stilling til behandling af outliers i datasættet. Det kan på baggrund af datasættet konstateres, at der forekommer ekstreme værdier af responsvariablen realiseret EPS. I figur 3.6 illustreres fordelingen af realiseret EPS værdier, og det fremgår hertil, at den højeste og laveste realiseret EPS værdi observeret i datasættet er henholdsvis 3,82 og -10,8, hvilket virker usandsynligt højt. Ifølge Barth & Kallapur (1996) kan outliers medføre biased estimater, hvilket alt andet lige påvirker generaliserbarheden af resultater. I litteraturen behandles outliers hovedsageligt ved én af følgende metoder, hvor outliers; beholdes i datasættet, winsorizes, eller ekskluderes fra datasættet (Ghosh & Vogt, 2012). Endvidere er der i litteraturen indenfor indtjeningsforecast ikke konsensus i forhold til hvilken af nævnte metoder, der benyttes til behandling af outliers. Xinyue et al. (2021) og van Binsbergen et al. (2020) jævnfør afsnit 2.2.3 benytter winzorization for at minimere outliers indflydelse på resultaterne. Ydermere vælger Elend et al. (2020) at fjerne observationer, hvor realiseret EPS udgør 1% kvartil i hver hale. Adskillelige studier oplyser ikke omkring behandlingen af outliers, hvortil det må antages at der ikke korrigeres for outliers i datagrundlaget.

Da vi i afhandlingen ønsker at korrigere for outliers i responsvariablen realiseret EPS, er det fundet hensigtsmæssigt at følge Elend et al. (2020) tilgang, og fjerne de observationer hvor realiseret EPS udgør de henholdsvis 1% laveste og højeste værdier. Såfremt winzoriation på baggrund af forecast fejl blev benyttet ville dette endvidere resultere i, at valide observationer fjernes fra datasættet, hvorfor anvendelsen af denne metode ikke er fundet hensigtsmæssig i afhandlingen. Dette udgør ifølge

Figur 3.6: Fordeling af realiseret EPS



Note: Figuren illustrerer fordelingen af realiseret EPS værdier i datasættet.

Tukey (1960) et reelt problem for statistiske test, da stikprøven er manipuleret i en sådan grad, at datagrundlaget ikke længere kan betegnes som en stikprøve

4 Teoretisk grundlag

Formålet med dette kapitel er at etablere en teoretisk forståelsesramme for anvendelsen af supervised machine learning algoritmer. I kapitlet vil det teoretiske grundlag for supervised learning præsenteres og der vil yderligere foretages en beskrivelse af de mest populære supervised machine learning algoritmer. Afslutningsvis vil forskellige performance mål for machine learning algoritmers forecast præcision evalueres med udgangspunkt i afhandlingens problemstilling.

Key takeaways: Blandt de mest populære machine learning algoritmer er lineære modeller, beslutningstræer, ensemble modeller samt Support Vector Machine. Det mest optimale algoritme valg afhænger af den pågældende problemstilling, hvortil en høj generaliserbarhed af algoritmens resultater er essentiel. Mean absolute error (MAE) og skill-score opgjort på baggrund af MAE blev i kapitlet identificeret som de mest hensigtsmæssige performance mål til evaluering af machine learning modellens performance i afhandlingen.

4.1 Supervised Learning

Machine learning defineres af Murphy (2012) som et sæt af metoder, der automatisk kan identificere mønstre og tendenser i data, og på baggrund heraf kan forudsige fremtidige værdier.Med andre ord handler machine learning om at generere viden på baggrund af data og er et forskningsfelt, der går på tværs af både statistik, kunstig intelligens og computervidenskab.Der eksisterer overordnet to forskellige kategorier af machine learning, henholdsvis supervised- og unsupervised learning. Supervised learning udgør den mest anvendte samt mest succesfulde type af machine learning, og det er endvidere denne type machine learning, der vil benyttes i afhandlingen. Unsupervised learning vil derfor ikke beskrives yderligere i afhandlingen. Supervised learning benyttes til at forudsige bestemte outputværdier på baggrund af udvalgte inputvariable. I supervised learning oplyser brugeren både inputvariable og outputværdien for hver enkelt observation i træningsdatasættet, og det er heraf betegnelse supervised learning stammer fra, da brugeren superviserer algoritmen. Endvidere oplyses kun inputvariable i testsættet, hvortil algoritmen baserer sit endelige forecast på baggrund af den sammenhæng mellem inputvariable og outputværdier identificeret i træningsdatasættet (Müller & Guido, 2016).

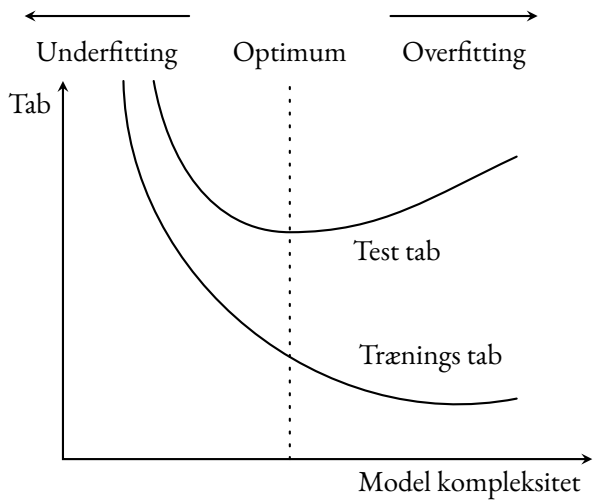
Supervised learning kan overordnet inddeles i de to problemstillingstyper; klassificering og regression. Målet i klassificeringsproblemstillinger er at forudsige én klasse på baggrund af en prædefineret liste af klasser (Müller & Guido, 2016). Et eksempel er studiet af Ou & Penman (1989) jævnfør afsnit 2.2.3, der ved brug af supervised learning forsøger at forudsige om et selskabs indtjening henholdsvis stiger eller falder det næstkommende kvartal. Modsat klassificering er målet i regressionsproblemstillinger at forudsige kontinuert værdier. Da vi i afhandlingen ønsker at forudsige selskabers EPS ved brug af machine learning, vil dette dermed kategoriseres som en regressionsproblemstilling.

4.1.1 Generaliserbarhed

Formålet med supervised learning er at konstruere en machine learning model, der på baggrund af træningsdata, er i stand til at forecaste på baggrund af ny og uset data. Såfremt en model er i stand til at generere præcise forecast på uset data, har modellen en høj generaliserbarhed på baggrund af træningsdata. Det ønskes endvidere at konstruere en model, der er i stand til at generalisere med så høj en forecast præcision som muligt. Ved optimering af modellers generaliserbarhed er det essentielt at undgå både underfitting og overfitting (Müller & Guido, 2016, Kapitel 2).

Antallet af inputvariable samt regler opstillet i en machine learning model er med til at bestemme kompleksiteten af modellen. En model bestående af adskillelige inputvariable og regler vil alt andet lige have en høj kompleksitet, og modsat vil en simpel model bestående af få inputvariable og regler have en lav kompleksitet. Konstrueres en machine learning model med en for høj kompleksitet betegnes dette som overfitting. Overfitting opstår, når modellen er for specifikt tilpasset til egenskaberne i træningsdata og hertil fokuserer for meget på individuelle datapunkter i træningssættet, men ikke er i stand til at generalisere på testdata. Modsat overfitting opstår underfitting, når modellen er for simpel til at opfange strukturelle mønstre i data, og modellen vil hertil opnå en lav forecast præcision på både træningsdata og testdata (Müller & Guido, 2016, Kapitel 2). Sammenhængen mellem model kompleksitet og forecast præcision er illustreret i figur 4.1.

Figur 4.1: Kompleksitet Vs. forecast præcision



Note: Figuren illustrerer sammenhængen mellem model kompleksitet og model tab for trænings- og testsættet.

4.2 Supervised machine learning algoritmer

Der findes adskillelige supervised machine learning algoritmer, der kan appliceres i regressionsproblemstillinger. Det er endvidere ikke muligt at fastlægge hvilken machine learning algoritme, der er bedst til en given problemstilling på forhånd, hvorfor dette heller ikke er tilfældet i afhandlingen. I følgende afsnit vil nogle af de mest udbredte typer af supervised machine learning modeller beskrives, herunder lineære, beslutningstræ, ensemble samt Support Vector Machine. Ydermere vil fordele og ulemper ved de respektive machine learning modeller evalueres.

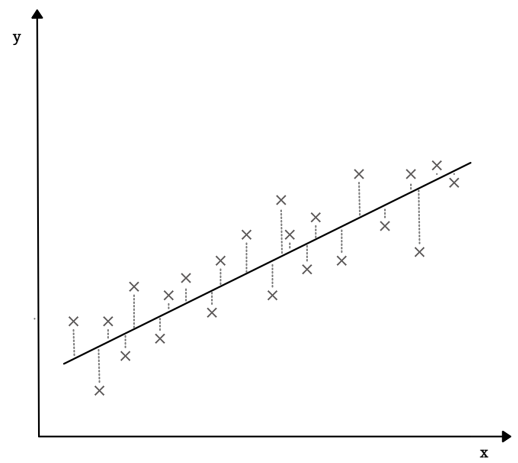
4.2.1 Lineære modeller

Lineære modeller benyttes ofte blandt praktikere og er beskrevet omfattende i litteraturen de seneste årtier. I lineære modeller baseres forecast på en lineær funktion af de valgte inputvariable og lineære modeller kan udtrykkes ved den generelle forecast formel jævnfør Scikit Learn (2020a):

$$\hat{y}(w, x) = w_0 + w_1 x_1 + ... + w_p x_p \tag{4.1}$$

I formlen angiver $\widehat{y}$ den estimerede værdi, w_0 angiver skæringen, $w_1 ... w_p$ udgør koefficienterne, mens $x_1 \ldots x_p$ angiver værdien af de uafhængige variable. I figur 4.2 illustreres et eksempel på en lineær model med én uafhængig variabel.

Figur 4.2: Lineær regression eksempel



Note: Figuren illustrerer et eksempel på en lineær regression.

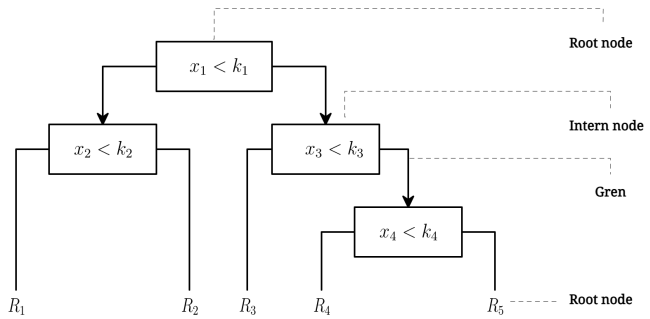
Lineære modellers antagelse om, at responsvariablen udgør en lineær kombination de respektive inputvariable kan virke begrænset, og som figur 4.2 antyder, betyde at modellen ikke at opfanger mønstre og tendenser i det underliggende data. Lineære modeller har dog vist sig at performe særdeles godt i tilfælde, hvor datasættet består af mange inputvariable. Såfremt datasættet består af flere inputvariable end observationer har lineære modeller vist sig at estimere responsvariablen med en høj forecast præcision (Müller & Guido, 2016, Kapitel 2).

Der findes mange forskellige lineære modeller, der kan benyttes i regressionsproblemstillinger, herunder Ordinary Least Squares, Ridge Regression og Lasso Regression. Forskellen mellem disse modeller består hovedsageligt i, hvordan de respektive modellers $w_1 \ldots w_p$ og w_0 parametre lærer fra træningsdata (Müller & Guido, 2016, Kapitel 2).

4.2.2 Beslutningstræer

Beslutningstræ algoritmer blev introduceret af Breiman et al. (1984) og er sidenhen blevet en af de mest populære machine learning algoritmer blandt praktikere og akademikere. Jævnfør afsnit 2.2.3 gør blandt andet van Binsbergen et al. (2020) og Hunt et al. (2019) brug af en beslutningstræ model i forbindelse med forecasting af EPS. Grundlæggende for beslutningstræ algoritmer er, at de lærer et hierarki af if-else statements, og forsøger på baggrund heraf at estimere responsvariablen. Beslutningstræ algoritmer benytter endvidere et sæt af binære regler til at foretage det endelige estimat, hvilket et illustreret i figur 4.3. Beslutningstræer består af en *root node* der udgør den øverste node og *interne noder* der repræsenterer grænseværdier opstillet for de respektive inputvariable også kaldet tests. Ydermere består beslutningstræer af *grene,* der angiver udfaldet de respektive tests, samt *leaf noder,* der angiver den estimerede værdi baseret på et gennemsnit af responsvariable værdier i træningssættet, givet de samme grænseværdier (Boehmke & Greenwell, 2020, Kapitel 9).

Figur 4.3: Beslutningstræ eksempel



Note: Figuren illustrerer et eksempel på, hvordan et beslutningstræ løser en regressionsproblemstilling.

Ifølge Boehmke & Greenwell (2020, Kapitel 9) bestemmes grænseværdierne for hver intern node ved, at identificere den grænseværdi, hvor summen af de kvadrerede residualer (SSE) minimeres jævnfør formel 4.2.

$$SSE = \sum_{i \in R_1} (y_i - c_1)^2 + \sum_{i \in R_2} (y_i - c_2)^2 \tag{4.2}$$

Såfremt en beslutningstræ model består af mere end én inputvariabel gentages denne proces for alle inputvariable, for at identificere de optimale grænseværdier. Inputvariablen med den laveste SSE vælges hertil som root noden. Efterfølgende identificeres den inputvariabel med lavest SSE sammenlignet med de resterende inputvariable, og denne vælges hertil som en ny intern node i forlængelse af root nodens gren. Denne proces gentages indtil alle inputvariable i træningssættet indgår i beslutningstræet, samt at hver *leaf node* kun indeholder en enkelt estimeret værdi. Når

beslutningstræerne er konstrueret, kan responsvariablen i testsættet estimeres ved at starte fra root noden og evaluere hver interne nodes i beslutningstræet, indtil én *leaf node* nås. *Leaf noden* angiver hertil det endelige estimat. Ved modellering af beslutningstræer udgør størrelsen af træet en vigtig faktor for modellens forecast præcision. Beslutningstræer, der består af mange noder, har en tendens til overfitting, hvilket resulterer i en lav forecast præcision på testdata. Endvidere vil beslutningstræer, der kun består af enkelte noder ikke opfange detaljegraden i det underliggende data, hvilket kan resultere i underfitting.

Overordnet set har beslutningstræ modeller den fordel sammenlignet med andre algoritmer, at de er simple samt relativt nemme at fortolke for folk med begrænset statistisk viden. Ydermere påvirkes beslutningstræ modeller ikke af forskellige typer af inputdata (e.g. kategoriske, diskrete, kontinuerte etc.), og forudsætter derfor ikke skalering i form af normalisering eller standardisering, modsat andre typer af algoritmer. Beslutningstræ modeller har endvidere vist sig at være særligt anvendelige, såfremt der benyttes inputvariable opgjort på forskellige skalaer, og ydermere ved en kombination af kategoriske og kontinuerte inputvariable. Den væsentligste ulempe ved beslutningstræ modeller er, at disse ofte har en tendens til at overfitte træningsdata. Ydermere udgør en høj varians ligeledes en udfordring, da mindre ændringer i data kan medføre signifikante ændringer i opbygningen af beslutningstræet, hvilket gør det vanskeligt for beslutningstræer at estimere data præget af støj (Müller & Guido, 2016, Kapitel 2; Boehmke & Greenwell, 2020, Kapitel 9).

4.2.3 Ensemble learning

Ensemble learning går grundlæggende ud på at konstruere en model ved at kombinere styrker fra forskellige machine learning modeller. Ensemble learning benyttes hovedsageligt for at forbedre forecast præcision samt at reducere sandsynligheden for et dårligt modelvalg. Hastie et al. (2009, Kapitel 16) beskriver ensemble learning som processen hvori, der udvikles en samling af machine learning modeller kaldet *base learners* på baggrund af træningsdata, og at disse herefter kombineres til en samlet forecast model. Ifølge Murphy (2012, s. 580) kan vægtningen af de respektive base learners betegnes som:

$$f(y \mid \mathbf{x}, \boldsymbol{\pi}) = \sum_{m \in \mathcal{M}} w_m f_m(y \mid \mathbf{x}) \tag{4.3}$$

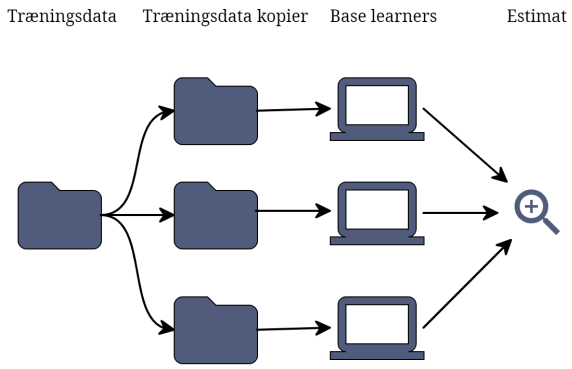
Hvor w_m angiver vægtning mens f_m angiver de respektive base learners.

Ensemble metoder kan overordnet inddeles i tre overordnet teknikker; bagging, boosting og stacking. I det følgende afsnit vil de væsentligste karakteristika for hver enkelt ensemble metode gennemgås.

Bagging

Bootstrap aggregering også kaldet bagging, er en af de første ensemble baserede teknikker, og har til formål at øge stabiliteten samt forbedre forecast præcision af algoritmer. I litteraturen har bagging opnået lovende resultater og teknikken kan endvidere benyttes på alle typer af machine learning algoritmer, men anvendes oftest på beslutningstræ modeller. I bagging genereres flere versioner af forecast, hvilke benyttes til at opnå et aggregeret forecast. I figur 4.4 illustreres grundprincipperne bag bagging.

Figur 4.4: Bagging metoden



Note: Figuren illustrerer bagging metoden.

Bagging algoritmen foretager bootstrapping[9] kopier af træningsdataet, hvorefter individuelle base learners appliceres på hver enkelt bootstrapping kopi. Endvidere foretager de individuelle base learners et forecast på baggrund af de respektive bootstrapping kopier. Det aggregerede forecast opnås hertil ved at tage det gennemsnitlige forecast af de respektive base learners forecast (Breiman, 1996b). Aggregering af de individuelle base learners forecast i bagging teknikken medvirker til en reducering af variansen fra individuelle base learner. Overordnet set øger bagging forecast præcisionen for modeller med høj varians, der oplever markante ændringer i forecast som effekt af mindre ændringer i træningsdata (Dietterich, 2000). Modsat er bagging mindre hensigtsmæssigt for mere stabile base learners, da det ikke er muligt at forbedre forecast i samme omfang, grundet den lavere varians i forecast. Den væsentligste ulemperne ved bagging er endvidere, at bagging teknikken er tidskrævende og alt andet lige vanskeliggøre fortolkning den endelige model. Der findes endvidere adskillelige machine learning algoritmer, der benytter bagging teknikken, herunder Random Forest algoritmen, der udgør en af de mest populære machine learning algoritmer.

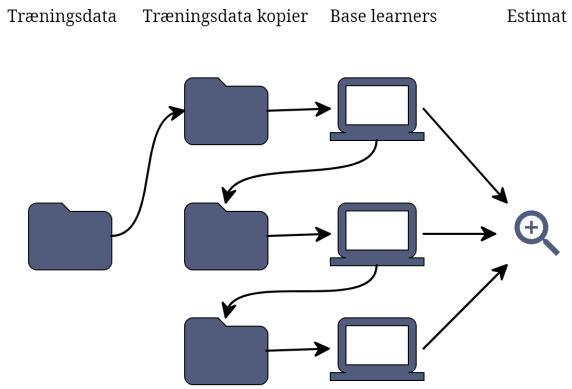
Boosting

Modsat bagging adskiller boosting metoden sig ved at have stor anvendelighed ved modeller med en høj grad af bias og lav varians. Boosting kan ligesom bagging benyttes ved forskellige typer af algoritmer, men er oftest mest effektivt i beslutningstræmodeller, hvorfor dette vil være udgangspunktet for resten af afsnittet.

Den grundlæggende ide bag boosting er at tilføje nye modeller til en samlet gruppe af machine learning algoritmer sekventielt. Boosting forsøger at minimere bias og varians i modellen, ved at tage udgangspunkt i en simpel model (eksempelvis beslutningstræ med få split) og booster hertil performance, ved at tilføje nye træer. Hvert enkelt træ der tilføjes, forsøger at forbedre punkter, hvor det foregående træ foretog de største fejl. Endvidere optimeres forecast fejl ved at modellen fokusere på træning af de observationer, hvor det foregående træ foretog de største forecast fejl (Boehmke & Greenwell, 2020, Kapitel 12). Et eksempel på den sekventielle ensemble teknik boosting er illustreret i figur 4.5.

[9] Bootstrapping er en resampling metode, der bruger tilfældig prøveudtagning.

Figur 4.5: Boosting metoden



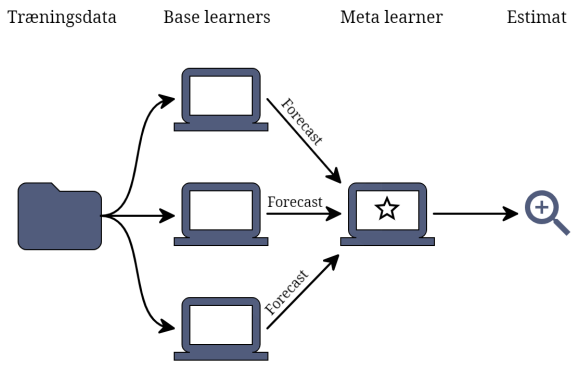
Note: Figuren illustrerer bagging metoden.

Boosting algoritmer har til trods for at være mindre intuitive og mere beregningstunge, ofte vist sig at outperforme andre machine learning modeller. Endvidere er AdaBoost, XGBoost og LightGMB nogle de mest populære boosting modeller og er anerkendte af praktikere for deres suveræne performance (Boehmke & Greenwell, 2020, Kapitel 12).

Stacking

Stacking blev oprindeligt introduceret af Wolpert (1992), og blev efterfølgende videreudviklet af Breiman (1996a), der skabte den mere moderne form for stacking, der benyttes i dag.I stacking trænes *base learners* først, hvorefter en meta algoritme, også betegnet *super learner,* foretager det endelige forecast baseret på de respektive base learners forecast(Boehmke & Greenwell, 2020, Kapitel 15).

Figur 4.6: Stacking metoden



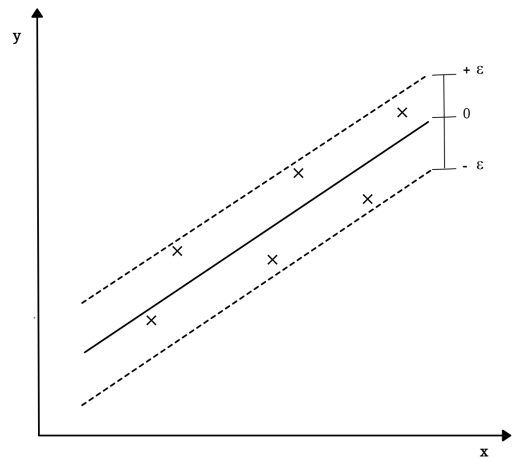
Note: Figuren illustrerer stacking metoden.

I figur 4.6 illustreres grundprincippet i stacking teknikken. Indledningsvis trænes *base learners* med udgangspunkt i træningsdata, hvorefter hver enkelt base learner foretager et forecast. På baggrund af de respektive forecast trænes en meta algoritmen, hvilket gør det muligt for meta algoritmen at lære af de respektive base learners forecast fejl. Meta algoritmen forsøger hertil at korrigere de respektive base learners forecast fejl og foretager hertil et endeligt forecast. Stacking teknikken har endvidere vist lovende resultater og har en tendens til at outperforme individuelle machine learning algoritmer (Boehmke & Greenwell, 2020, Kapitel 15).

4.2.4 Support Vector Machine

Support Vector Machine (SVM) algoritmen blev introduceret af Cortes og Vapnik (1995) og udgjorde oprindeligt en lineær klassificeringsmetode til klassifikationsproblemstillinger. SVM er sidenhen udviklet til yderligere at understøtte regressionsproblemstillinger, hvilket betegnes Support Vector Regression (SVR). Grundprincippet i SVM er, at algoritmen forsøger at konstruere et lineært hyperplan, der kan adskille to udfald fra hinanden med højest mulig præcision. SVR-algoritmen adskiller sig fra SVM, ved at inkorporere modelparametre ϵ, der angiver en fejl margin, hvor fejl inden for denne grænse ignoreres i optimeringsprocessen. I figur 4.7 illustreres grundpricinperne bag SVM modeller.

Figur 4.7: Support Vector Regression (SVR) eksempel



Note: I figuren illustreres et eksempel på en SVR model. I eksemplet angiver de stiplede linjer grænseværdierne for fejlmarginen.

Som illustreret i figur 4.7 forsøger SVR algoritmen at konstruere et hyperplan, der opfanger flest mulige observationer inden for ε grænserne. Endvidere tager ovenstående illustration udgangspunkt i et todimensionelt rum, hvor mere end én inputvariabel vil betyde flere dimensioner med samme grundlæggende optimerings tankegang. For at SVR kan benyttes såfremt der ikke forekommer et lineært forhold mellem input- og responsvariabel anvendes kernel funktioner. Tankegangen bag kernel funktioner er at øge dimensionsrummet yderligere indtil, der kan tegnes en lineær sammenhæng. Polynomium, RBF og sigmoid udgør nogle af de mest populære kernels i SVR. Nogle af de primære fordele ved SVR modeller er, at de er særligt anvendelige ved mange inputvariable samt

stadig formår at opnå en god performance i tilfælde hvor antallet af inputvariable er større end antallet af observationer. Den væsentligste ulempe ved SVR er dog at de kan være langsomme at træne sammenlignet med andre typer af algoritmer (Scikit-learn, 2020b).

4.3 Evaluering af model performance

I følgende afsnit vil forskellige mål til evaluering af machine learning modeller samt analytikeres forecast præcision uddybes og diskuteres med udgangspunkt i afhandlingens problemstilling. I litteraturen er adskillelige metoder til måling af forecast præcision identificeret med udgangspunkt i forecast fejl, herunder skala-afhængige, procentuelle og skalerede performance mål. Indenfor brugen af tidsseriedata er der i litteraturen fremstillet forskellige anbefalinger af, hvilke metoder der er mest hensigtsmæssige som performance mål. Forskning inden for indtjeningsforecast har ligeledes benyttet forskellige metoder til vurdering af modeller forecast præcision. I studiet af Ball & Ghysels (2017) jævnfør afsnit 2.2.2, benyttes et procentuelt performance mål i form af median absolutte error ratio (MABER) til evaluering af forecast præcisionen. Ydermere benytter Elend et al., (2020) mean squared error (MSE) som performance mål til evaluering af Deep Learning modellers forecast præcision i forbindelse med estimering af selskabers fremtidig EPS.

Manglen på et standard performance mål til evaluering af forecast præcision vanskeliggør endvidere sammenligning af resultater på tværs af studier. Ifølge Hyndman & Koehler (2006) er mange performance mål anvendt i litteraturen ikke repræsentative og kan i værste fald producere misvisende resultater, hvilket understreger vigtigheden af en korrekt målemetode.

4.3.1 Performance på samme datagrundlag

Skala-afhængige forecast fejl er kendetegnet ved at tage udgangspunkt i same skala som datasættet. Denne måde at opgøre forecast fejl er særligt anvendelig ved sammenligning af forecast præcision på tværs af forskellige forecast metoder baseret på det samme datasæt, men bør ikke anvendes til sammenligning af forecast præcision på tværs af datasæt, med forskellige skalaer (Hyndman & Koehler, 2006). Mean absolute error (MAE), mean squared error (MSE) og root mean squared error (RMSE) udgør nogle af de mest anvendte skala-afhængige performance mål (Hyndman & Athanasopoulos, 2021). Historisk set har RMSE og MSE været særlig populære grundet deres teoretiske relevans i statistisk modellering. Disse performance mål er dog mere følsomme overfor outliers end MAE, hvorfor visse forskere (eksempelvis: J. Armstrong, 2001, Kapitel 14) anbefaler at disse ikke benyttes til evaluering af forecast præcision (Hyndman & Koehler, 2006).

Da afhandlingen ønsker at sammenligne performance mellem henholdsvis en machine learning model og analytikere baseret på et identisk datagrundlag, og dermed samme skala, er det fundet hensigtsmæssigt at anvende både MAE og RMSE som performance mål i afhandlingen. Jævnfør Boehmke & Greenwell (2020, Kapitel 2) beregnes MAE ved følgende formel 4.4:

$$\text{MAE} = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j| \tag{4.4}$$

Hvor n angiver antallet af observation, y_i udgør den sande EPS værdi og $\widehat{y}_i$ angiver den estimerede EPS.

Ydermere beregnes performance målet RMSE, ved anvendelse af formel 4.6:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_j - \hat{y}_j)^2 \tag{4.5}$$

$$\text{RMSE} = \sqrt{\text{MSE}} \tag{4.6}$$

Som illustreret i formel 4.6 udgør RMSE kvadratroden af MSE. I formlen angiver n antallet af observationer i datasættet, y_i angiver den sande værdi af EPS, mens $\widehat{y}_i$ udgør den estimerede EPS. Outliers store indflydelse på RMSE, som blandt andet er kritiseret af J. Armstrong (2001), vil performance målet kun benyttes overfladisk i afhandlingen, til identificering af eventuelle outliers.

4.3.2 Performance ved forskelligt datagrundlag

I afhandlingen ønsker vi jævnfør arbejdsspørgsmål 1 at undersøge forskelle i en machine learning models performance ved forskellige forecast horisonter samt på tværs af industrier. For at undersøge dette er det dermed essentielt sammenligne forecast, der baserer sig på forskellige datagrundlag. Da skala-afhængige performance mål, herunder MAE og RMSE ikke bør anvendes til sammenligning af data på baggrund af forskellige datagrundlag, er det hertil essentielt at inddrage andre performance mål (Hyndman & Athanasopoulos, 2021). Procentuelle performance mål kan hertil benyttes, da denne type af performance mål ikke tager udgangspunkt i datagrundlagets skala, hvilket muliggør sammenligning af performance ved forskellige datagrundlag. Mean absolute percentage error (MAPE) udgør det mest anvendte procentuelle performance mål, men har dog en væsentlig ulempe i form af at være uendelig eller udefinerbar, såfremt den sande værdi er 0. Ydermere giver MAPE ekstreme resultater, såfremt den sande værdi er tæt på 0. Endvidere er negative residualer mere udslagsgivende end positive i MAPE, hvilket ligeledes udgør en væsentlig ulempe ved performance målet (Hyndman & Athanasopoulos, 2021). Da distributionen af realiseret EPS i afhandlingens datagrundlag centrerer sig omkring 0 som belyst i afsnit 3.2.9, er det ikke fundet hensigtsmæssigt at anvende MAPE som performance mål i afhandlingen.

Ifølge Hyndman & Athanasopoulos (2021, Kapitel 5) kan skala-fri og skalerede performance mål anvendes som et alternativ til procentuelle performance mål ved sammenligning af forecast præcision på tværs af datagrundlag. Hyndman & Athanasopoulos (2021, Kapitel 5) anbefaler hertil brugen af skill-score samt performance målet MASE. Som belyst i afsnit 2.2.1 vælger en stor del af studierne i litteraturen af sammenholde resultater med et naivt forecast. Af denne årsag er det fundet særligt hensigtsmæssigt at inddrage skills-score, da forecast præcisionen hertil evalueres med udgangspunkt i et naivt forecast. Endvidere er forecast præcisionen opgjort ved skill-score lettere at fortolke end forecast præcision opgjort ved MASE, da skill-scoren basere sig på testsættet og derved giver den eksakte outperformance modsat MASE, der approksimerer dette på baggrund af træningsdata. Skill-scoren vil hertil afspejle forecast præcisionen opgjort ved MAE relativt til det naive forecast (Hyndman & Athanasopoulos, 2021, Kapitel 5). Andre studier, herunder Elend et al. (2020) benytter ligeledes skill-score som mål for performance, men vælger hertil at opgøre skill-scoren

baseret på MSE. Da vi i afhandlingen ikke ønsker at tillægge særlig vægt på outliers, er det derfor fundet hensigtsmæssigt at opgøre skill-scoren med udgangspunkt i MAE. Ifølge (Hyndman & Athanasopoulos (2021, Kapitel 5) kan skill-score beregnes på følgende måde:

$$\mathrm{SS}_{\mathrm{MAE}} = 1 - \frac{\mathrm{MAE}(m)}{\mathrm{MAE}(\mathrm{naiv})} \tag{4.7}$$

I formel 4.7 angiver *MAE (m)* modellens MAE, mens *MAE (naiv)*udgør det naive forecasts MAE. Såfremt en skill-score på mere end 0 opnås vidner dette om en forecast præcision, der er bedre end den naive model, mens en skill-score på mindre end 0 vidner om en dårligere forecast præcision. Skill-scoren vil anvendes som performancemål i forbindelse med sammenligning af performance på tværs af datagrundlag, herunder særligt på tværs af industrier og tidshorisonter.

5 Modellering

I følgende kapitel vil de væsentligste aspekter i forbindelse med valg af machine learning algoritme samt optimering af modellens performance gennemgås. Indledningsvis vil DataRobot platformen, der benyttes eksplorativt til udvælgelse af den mest optimale machine learning algorithme beskrives. Dernæst foretages en opdeling af trænings- og testsæt samt et endeligt valg af den mest optimale machine learning algoritme og dertilhørende inputvariable samt model parametre.

Key takeaways: På baggrund af cross validation resultater blev Catboost algoritmen identificeret som det mest optimale algoritme valg, givet afhandlingens datagrundlag. På baggrund af feature engineering og feature selektion samt hyperparameter tuning, blev de mest optimale inputvariable samt model parametre identificeret, hvilke danner grundlaget for resultaterne i kapitel 6.

5.1 DataRobot

I tilstræbelsen på at opnå det bedste analysemæssige grundlag for afhandlingen benyttes DataRobots machine learning platform eksplorativt, for at udforske anvendeligheden af forskellige machine learning algoritmer på afhandlingens problemstilling. DataRobot er en automatiseret machine learning platform, der ved at kombinere machine learning teknologi, øger hastigheden og præcisionen af forecast modeller og yderligere muliggør kombinationen af tidsserie, geografisk information og tabeldata (DataRobot, 2021). DataRobots platform anvendes af adskillelige multinationale organisationer, herunder Lenovo og Deloitte, og blev i år 2019 anerkendt af Forbes, som en af USA's mest lovende AI virksomheder, med en sjette plads på Forbes AI 50 liste[10] (D'Onfro, 2019).

Til trods for DataRobot platformens imponerende egenskaber benyttes DataRobot udelukkende eksplorativt, grundet de begrænsede muligheder for at opnå en dybdegående indsigt i forecast resultater. Ydermere har DataRobots manglende transparens i behandling og klargørelse af data i forbindelse med applicering machine learning algoritmer, herunder transformering at kategoriske variable og normalisering af data. Endvidere er kun et begrænset antal machine learning algoritmerne tilgængelige på DataRobots platform og platformen giver dermed ikke mulighed for at inkludere nyligt udviklede algoritmer. Som afsnit 2.2.3 er det ikke muligt på forhånd at fastlægge hvilken machine learning algoritme, der er bedst til en given problemstilling. Fordelen ved at benytte DataRobots platform eksplorativt er hertil, at dette giver mulighed for at få et overblik over forskellige machine learning algoritmers performance baseret på afhandlingens datagrundlag. Som belyst i afsnit 4.2 kan adskillelige machine learning algoritmer benyttes til at estimere selskabers EPS, såsom beslutningstræmodeller, SVM og boosting modeller. Anvendelsen af DataRobots platform danner endvidere et solidt grundlag for udvælgelsen af en endelig machine learning algoritme til

[10] "AI 50: America's Most Promising Artificial Intelligence Companies"

afhandlingens fremsatte problemstilling, da der omtrent blev eksperimenteret med 100 forskellige algoritmer på datagrundlaget.

Udvælgelsen af machine learning algoritmer tager under normale omstændigheder udgangspunkt i en selektiv udvælgelse af modeller, hvorudfra forecast præcisionen evalueres og en endelig model udvælges. Såfremt denne metode blev implementeret i afhandlingen, ville konklusioner fremstillet i afhandlingen bære præg af selection bias. Denne problematik minimeres ved anvendelsen af DataRobot, da platformen sikrer et mere validt udvælgelsesgrundlag.

5.2 Opdeling af data i trænings- og testsæt

I forbindelse med modellering af machine learning algoritmer er det normal praksis at foretage en opdeling af datagrundlaget i henholdsvis trænings-, validerings- og testsæt. Træningssættet består af henholdsvis inputvariable samt responsvariablen, hvilket machine learning modellen benytter til at lære af, ved at identificere generelle mønstre og tendenser. Valideringssættet benyttes til at evaluere den trænede models performance, inden den endelige machine learning model testes på uset data. Valideringssættet er dermed med til at sikre, at der ikke forekommer reverse engineering i modelleringsprocessen. Ydermere giver valideringssættet en indikation af, hvorvidt machine learning modellen overfitter[11] på træningssættet. Testsættet benyttes først i den afsluttende fase, hvori den endelige machine learning models forecast præcision testes på uset data. Det er endvidere testsættet, der benyttes til at evaluere den endelige models forecast præcision, og resultaterne heraf vil endvidere sammenlignes med analytikeres forecast præcision (Müller & Guido, 2016, Kapitel 1).

Under normale omstændigheder blandes det samlede datasæt tilfældigt, hvorefter en opdeling af datasættet i henholdsvis trænings-, validerings- og testsæt foretages, for at undgå en systematisk opdeling af datasættet. I tilfælde af at datasættet blandes før opdelingen, vil den kronologiske rækkefølge i datasættet ignoreres, hvilket udgør et problem ved tidsseriedata (Kraus & Feuerriegel, 2017). Såfremt der ikke tages højde for den kronologiske rækkefølge i datasættet, ville look-ahead bias forekomme, da trænings- og valideringssættet fejlagtigt ville drage fordel af information, der ellers kun ville have været tilgængelig ex-post. Endvidere ville look-ahead bias andet lige resultere i en højere forecast præcision i modellen, men dette ville dog ikke afspejle et virkelighedsscenarie ved efterfølgende implementering af modellen. For at undgå look-ahead bias foretages opdelingen i henholdsvis trænings, validerings og testsæt i afhandlingen på baggrund af et kronologisk split i følgende rækkefølge; træning, validering og test. Den kronologiske opdeling af datagrundlaget fremgår af tabel 5.1.

5.3 Cross validation

I regressionsproblemstillinger er det som belyst i afsnit **??**, essentielt at tage stilling til generaliserbarheden af den givne model. Med andre ord er modellens evne til at opnå en høj forecast præcision ved forecast af nyt uset data essentielt. Valideringssættet giver endvidere ikke en endegyldig indikation af, hvordan modellen vil performer på uset data, men er dog indikativ for, om der

[11] Læs mere herom i afsnit 4.1.1

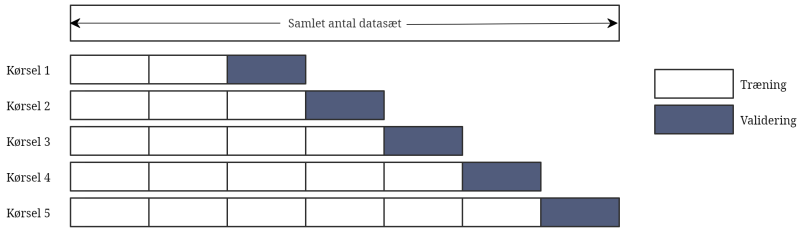
Tabel 5.1: Opdeling af datasæt

Data	Træning	Validering	Test
Periode	30/06/1998 - 30/06/2016	30/9/2016 - 30/6/2018	30/9/2018 - 30/6/2020
Virksomheder	3.536	1.968	1.758
% af datasæt	80%	10%	10%

Note: Tabellen angiver andelen af observationer i datasættet fodelt på trænings-, validerings og testsættet.

forekommer henholdsvis under- eller overfitting i modellen (Gupta, 2017). For at validere robustheden af machine learning modeller kan rolling origin cross validation (ROCV) benyttes, da denne teknik benyttes specifikt til at evaluere robustheden af tidsserie forecast modeller (Hyndman & Athanasopoulos, 2021). I ROCV inddeles forecast modellen i forskellige sektioner af data. Hovedformålet er hertil at træne og teste forecast modellen på initiale split, hvorefter hver sektion af træning og test forskydes en periode frem. Forskydningen fortsættes indtil al tilgængelige data er udtømt. ROCV processen er illustreret i figur 5.1.

Figur 5.1: Rolling origin cross validation eksempel

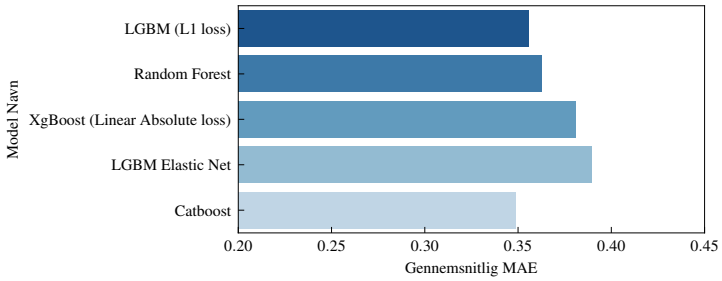


Note: Figuren illustrerer et eksempel på Rolling origin cross validation, hvor træningsperioden forlænges og testperioden forskydes for hver kørsel.

Den pågældende machine learning models forecast præcision beregnes på baggrund af et gennemsnit af alle valideringssæt, og resultater heraf indikerer modellens evne til at tilpasse nyt data. I tilstræbelsen på at udvælge machine learning modellen med den højeste generaliserbarhed, benyttes et gennemsnit af cross-validation præcision på fem kørsler, som selekteringskriterie for valget af machine learning algoritme i afhandlingen. Dette betyder endvidere at evaluering af forskellige machine learning algoritmers forecast præcision på det endelige testsæt ikke testes, for at undgå bias i optimeringsprocessen.

På baggrund af DataRobot er det muligt at identificere ROCV performance ved implementering af forskellige machine learning algoritmer. I figur 5.2 fremgår ROCV performance for de fire algoritmer med den bedst performance ved en forecast horisont på 1 kvartal. ROCV er endvidere opgjort på baggrund af performance målet MAE. Forfatterne har ydermere valgt at inddrage en femte model i form af Catboost modellen, der er implementeret i Python, da denne model ikke er tilgængelig på Datarobots platform. Catboost algoritmen er anerkendt af machine learning praktikere for sin anvendelse på datasæt bestående af mange kategoriske variable, hvilket er tilfældet i afhandlingen (Dorogush et al., 2018). ROCV af Catboost modellen er implementeret ved brug af Scikit Learns *TimeSeriesSplit* funktion (Scikit Learn, 2020).

Figur 5.2: Cross validation Datarobot



Note: Figuren illustrerer de fem machine learning algoritmer med den laveste gennemsnitlige MAE.

Jævnfør figur 5.2 fremgår det, at Catboost algoritmen opnår den bedste ROCV performance, med en gennemsnitlig MAE på 0,349. Endvidere er Catboost modellen kun marginalt bedre end LightGBM (L1 loss), der har en MAE på 0,3557. Catboost opnår dog en markant bedre performance end de restrerende algoritmer, herunder XgBoost, Random Forest og LightGBM (Elastic net). På baggrund af ROCV performance fremgår det, at Catboost algoritmen er det mest optimale modelvalg givet datagrundlaget i afhandlingen. Catboost algoritmen vil på baggrund heraf danne fundamentet for den endelige machine learning model med henblik på besvarelse af afhandlingens fremstillede problemstilling. I den videre analysere vil Catboost algoritmen derfor refereres til som modellen analogt.

CatBoost udgør en gradient boosting machine learning model og er udviklet af den russiske søgemaksine Yandex og offentliggjort i år 2017. Catboost har vundet indpas hos machine learning praktikere grundet sin suveræne performance. I et studie af Prokhorenkova et al. (2017) fremgår det, at Catboost modellen outperformer tidligere nævnte boosting modeller jævnfør afsnit 4.2.3 ved forecast på baggrund af syv forskellige velkendte datasæt (Prokhorenkova et al., 2017). I litteraturen har Catboost algoritmen ligeledes udvist lovende resultater i forbindelse med finansiel data (Hancock & Khoshgoftaar, 2020). Catboost adskiller sig fra andre boosting modeller, herunder XGBoost og LightGBM, ved at understøtte brugen af kategoriske variable uden krav om dummy encoding (Dorogush et al., 2018). Ydermere adskiller Catboost algoritmens modelarkitektur sig fra andre algoritmer ved at implementere symmetriske beslutningstræer, hvilket medvirker til simplere trænning af modellen samt hurtigere træning ved understøttelse af GPU og yderligere fungerer beslutningstræstrukturen som en form for regularisering, hvilket er med til at reducere overfitting (Ershov, 2018).

Base model

I forbindelse med feature engineering og feature selektering samt hyperparameter tuning er det essentielt at etablere en base model af Catboost modellen med standard modelparametre, der endvidere vil udgøre baseline for forecast præcisionen og forsøges optimeret. Jævnfør bilag A.2 fremgår de inputvariable, der er anvendt i basemodellen. Den etablerede base model opnår en MAE på 0,40 ved en forecast horisont på henholdsvis et kvartal. Base modellen vil endvidere benyttes som baseline og ydermere forsøges optimeret gennem feature engineering og feature selektion.

5.4 Feature Engineering

Feature engineering udgør den essentielle machine learning proces, hvori domæne viden specifikt til den givne problemstilling benyttes til at transformere data, således at data repræsenteres på en mere hensigtsmæssig måde (Shekhar, 2018). Med andre ord dannes nye inputvariable på baggrund af eksisterende inputvariable, med henblik på at forbedre machine learning modellens forecast præcision (Ruder et al., 2019). I afhandlingen genereres nye inputvariable på baggrund af feature engineering, herunder tidslag af realiseret EPS og regnskabstal samt en yderligere opgørelse af handelsvolumen i beløb fremfor antal aktier.

Anvendelsen af tidslaggede inputvariable gør det endvidere muligt for machine learning modellen at identificere historisk udvikling i henholdsvis realiseret EPS samt regnskabstal for de respektive selskaber i datagrundlaget, hvilken alt andet lige forventes at forbedre modellens forecast præcision. Tidligere research inden for automatisering af indtjeningsforecast har ligeledes benyttet tidslaggede inputvariable i forecast modeller, heriblandt i studiet af Ball & Ghysels (2017) og Elend et al. (2020). I afhandlingen benyttes en initial tidslag på ti perioder, hvilket betyder at modellen får oplyst realiserede EPS værdier samt regnskabstal de seneste ti kvartaler for det enkelte selskab i hver observation. Anvendelsen af tidslaggede værdier kan medføre overfitting, såfremt for mange tidslags anvendes, da tidslag øger kompleksiteten af datasættet. I det efterfølgende afsnit 5.5 vil de valgte tidslag endvidere evalueres, med henblik på at benytte det antal tidslag, der har den mest optimale indflydelse på modellens forecast præcision.

Input variablen *VOL* angav i sin oprindelige form handelsvolumen opgjort i antal aktier. Dette format er indikativ for efterspørgslen af aktier i selskaberne i datagrundlaget, men danner ikke et sammenligneligt grundlag på tværs af virksomheder, hvorfor det alt andet lige er vanskeligt for machine learning modellen at afkode generelle mønstre og tendenser. Da aktiekurserne og antallet af aktier i selskaberne varierer bredt i datagrundlaget, er det hertil fundet hensigtsmæssigt også at opgøre handelsvolumen i beløb. Ydermere er det fundet hensigtsmæssigt at opgøre inputvariablen *DPS* i procent i stedet for beløb, da dette alt andet lige danner et sammenligneligt grundlag på tværs af selskaber. Udbytte procent er udregnet på baggrund af udbytte per aktie (DPS) og aktiekursen ved brug formel 5.1:

$$\text{Udbytte}\,\% = \frac{\text{DPS}}{\text{Aktiekurs}} \tag{5.1}$$

Afslutningsvis er datoen blevet dekomponeret til en år og måned variabel der muliggør at machine learning modellen kan aflæse tendenser i indtjening over tid, samt i specifikke kvartaler.

5.5 Feature selektion

Feature selektion udgør processen hvori insignifikante inputvariable identificeres og fjernes fra datagrundlaget, med henblik på at øge machine learning modellens forecast præcision. I feature selektion er formålet endvidere at mindske antallet af inputvariable, da dette reducerer kompleksiteten af datasættet, hvilket alt andet lige medfører hurtigere træning af modellen samt reducerer overfitting (Panda & Misra, 2021, Kapitel 2). I litteraturen beskrives forskellige feature selektions teknikker, der hovedsageligt kan inddeles i de tre kategorier; filter, wrapper og embedded

metoder (Zheng & Casari, 2018, Kapitel 2). Filter metoder benytter statistiske teknikker til at evaluere relationen mellem inputvariable og responsvariablen, og vurderer på baggrund heraf hvilke inputvariable, der anvendes i modellen. Modsat filter metoder, konstrueres der i wrapper metoder forskellige modeller med forskellige kombinationer af inputvariable, hvortil den kombination af inputvariable, der resulterer i den bedste model forecast præcision vælges (Kuhn & Johnson, 2013, s. 490). I Embedded metoder indgår feature selektion som en integreret del af machine learning modellen, og disse metoder anvender en kombination af henholdsvis filter og wrapper metoderne (Panda & Misra, 2021, Kapitel 2).

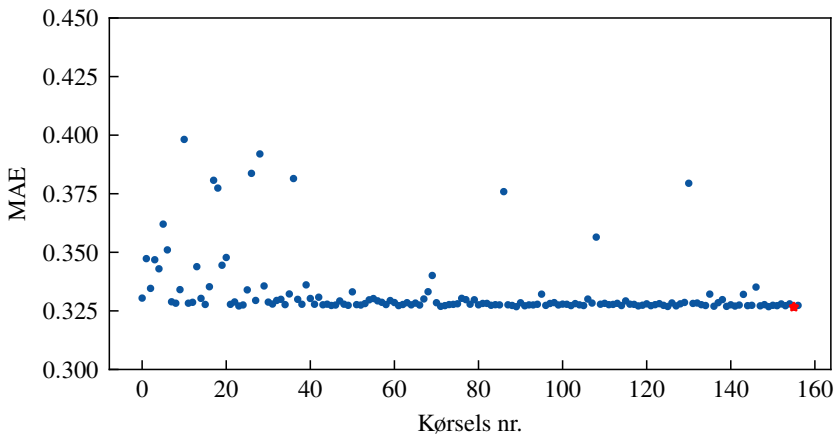
I afhandlingen benyttes wrapper metoden til feature selektion, da denne metode tager højde for kombination af inputvariable, hvilket betyder at inputvariable der er insignifikante individuelt men informative i en kombination ikke fjernes. Optimalt bør alle kombinationer af inputvariable og dertilhørende MAE identificeres, men grundet de talrige mulige kombinationer i datasættet, hvortil hver model trænes og evalueres på valideringssættet, er dette ikke muligt i afhandlingen. På baggrund heraf vil wrapper metoden i stedet tage udgangspunkt i forskellige kombinationer af henholdsvis tidslag, afkast, volumen samt prisvariables, da disse forventes at have en signifikant indflydelse på modellens forecast præcision. Det kan endvidere konstateres, at det mest optimale tidslag er 5 perioder. Ydermere fremgår det, at modellen opnår en bedre forecast præcision på valideringssættet såfremt *industry_return, VOL, volume_usd* samt *PRC* og *adj_close* fjernes. På baggrund af feature selektion er den mest optimale kombination af inputvariable dermed identificeret, og denne kombination vil udgøre det endelige inputgrundlag i modellen. I bilag A.3 fremgår en udtømmende liste over de mest optimal inputvariable.

5.6 Hyperparameter tuning

Machine learning algorithmer har adskillelige variable, der kan justeres, hvilke i praksis betegnes som hyperparametre. Hyperparametre udgør de parametre, hvis værdier benyttes til at kontrollere modellen læringsproces, eksempler herpå er regularisering, boostrap type samt valg af boosting metode. Processen hvori hyperparametre justeres med henblik på at optimere modellens performance kaldes i praksis hyperparameter tuning (Claesen & De Moor, 2015). I Hyperparameter tuning forbedres forecast præcision af den givne model, ved at den mest optimale model arkitektur vælges, givet det underliggende datagrundlag. I litteraturen beskrives adskillelige metoder til optimering af hyperparametre heriblandt Random Search, Grid Search, Bayesian optimering (O'REILLY, 2021).

I afhandlingen benyttes Optunas framework i Python til hyperparameter tuning, da denne modsat den klassiske metode Gridsearch ikke forsøger at prøve alle tænkelige kombinationer, men derimod løbende vælger de parameter værdier, der performer særligt godt på det specifikke datagrundlag (Akiba et al., 2019). Ydermere er Optuna nemt at implementere og kan samtidig integreres med Pythons syntaks.I Optuna frameworket defineres de modelparametre der ønskes optimeret, hvortil depth, bootstrap type og boosting type blandt andet er valgt. Valget af netop disse optimeringsparametre beror på CatBoost algoritmens officielle dokumentation og anbefalinger i forbindelse med hyperparameter tuning (CatBoost, 2021). De definerede optimeringsparametre benyttes i Optuna til at maksimere modellens cross-validation forecast præcision, ved at foretage så mange forskellige testkørsler som muligt på 16 timer. Implementeringen af Optuna blev foretaget på Microsoft Azure cloud platform og resulterede endvidere i 156 testkørsler.

Figur 5.3: Hyperparameter tuning - optimerings historik



Note: Figuren illustrerer den gennemsnitlige MAE ved forskellige kombinationer af hyperparametre. I figuren angiver den røde stjerne kørslen, med den mest optimale kombination af hyperparametre.

I figur 5.3 illustreres modellens performance opgjort for hver enkelt testkørsel, hvortil det fremgår at modellen med de mest optimale hyperparametre er testkørsel 155, der opnår en gennemsnitlig MAE på 0,3267. Ydermere kan det på baggrund af figur 5.3 konstateres, at hyperparametren *depth* har den største indflydelse på modellens forecast præcision. Da vi i afhandlinger ønsker at afspejle den mest optimale model til sammenligning med analytikeres forecast performance, vil modelarkitekturen i den endelige model på baggrund heraf, basere sig på hyperparametrene identificeret i testkørsel 155. De endelige model hyperparametre fremgår af tabel 5.2, og vil endvidere benyttes på alle analyser i afhandlingen.

Tabel 5.2: Catboost hyperparametre

Hyperparameter	Valgmuligheder	Optimeret værdi
Colsample by level	{0,01 - 0,1}	0,09
Depth	{1 - 12 }	8
Boosting type	{Ordered, Plain}	Plain
Bootstrap type	{Bayesian, Bernoulli, MVS}	MVS
Bagging temperature**	{0 - 10}	
Subsample**	{0,1 - 1}	
Max ctr complexity*		0
One hot max size*		5000

Note: Tabellen angiver hyperparamtre der optimeres.* angiver parameter værdier valgt på baggrund af Catboost dokumentation, for at sikre høj performance og hurtig træning. ** angiver ikke optimerede hyperparametre, da bootstrap typen ikke er Bayesian eller Bernoulli.

6 Resultater og analyse

I det følgende kapitel vil den endelige machine learning models resultater præsenteres og analyseres med henblik på at besvare den fremsatte problemstilling. Først vil modellens overordnede resultater analyseres, hvor forecast præcision vil sammenlignes med analytikerne. Dernæst vil det vurderes, hvorvidt det er hensigtsmæssigt at benytte industrispecifikke modeller fremfor en generel model, hvorefter den endelige model sammenholdes med forecast præcisionen blandt analytikere. Efterfølgende foretages en dekomponering af de respektive modellers resultater, baseret på henholdsvis virksomhedsstørrelse samt analytikerdækning og spredning i analytiker estimater, med henblik på at identificere, hvilke implikationer disse har for forecast præcisionen. Afslutningsvis undersøges modellens robusthed overfor ændringer i økonomiske konjunkturer samt længden af træningsperioden. I kapitlet vil de væsentligste inputvariable i de forskellige delanalyser løbende udforskes og analyseres.

Key takeaways: Den konstruerede machine learning model formår overordnet set at outperforme analytikere på tværs af alle forecast horisonter, og er endvidere statisk signifikant ved forecast horisonter på to og fire kvartaler. Machine learning modellen viser sig særligt anvendelig ved forecast af selskaber i industrierne; Healthcare, Energy og Chemicals & allied products. Ydermere er modellen særdeles anvendelig ved forecast af mindre selskaber, samt i tilfælde hvor der er en lav analytiker dækning og en høj spredning i analytiker estimater. Et robusthedstjek af modellen vidnede om, at modellens resultater er robust overfor ændringer i længden af trænings perioden samt ændringer i træning- og testperioden.

I tilstræbelsen på at opnå en høj generaliserbarhed af de fremlagte resultater, vil der løbende udarbejdes statistiske test, med henblik på at klarlægge, hvorvidt resultaterne er statisk signifikante. Den meste anerkendte og anvendte statistiske test til at sammenligne forecasts er testen udviklet af Diebold & Mariano (1995). Afhandlingen vil endvidere benytte en modificerede udgave af Diebold og Mariano testen (Harvey et al., 1997). Fordelen ved modifikationen af Diebold & Mariano testen er muligheden for at benytte ikke kvadrerede loss funktioner, hvilket muliggør brugen af absolutte fejl fremfor kvadrerede fejl. Dette er særligt relevant, da tidligere litteratur har identificeret, at en absolut loss function er mest repræsentativ for analytikeres loss function, samt at denne er mere robust overfor outliers i distributionen end andre loss funktioner (Abarbanell & Lehavy, 2003; Basu & Markov, 2004). Ydermere betyder modificeringen at Diebold & Mariano testen, at testen ikke længere beror på en antagelse om ingen bias i forecast fejl. Dette er særligt relevant for afhandlingen, da der som beskrevet i afsnit 2.1.2 historisk er identificeret et systematisk optimisme bias blandt analytikere.

Den modificerede Diebold og Mariano test er endvidere mere hensigtsmæssig på små og mellemstore stikprøver, grundet den oprindelige diebold og Mariano tests tendens til at forkaste nulhypotesen, selvom nulhypotesen er korrekt. Da statiske test i afhandlingen udføres på det testsættet, der kan betegnes som små stikprøver, er derfor fundet mest hensigtsmæssigt at benytte den modificerede

Diebold og Mariano test. Statistisk test foretaget i kapitlet vil derfor basere sig på nedenstående formel 6.2, der udgør den modificerede Diebold & Mariano test.

Jævnfør Diebold & Mariano (1995) og Harvey et al. (1997) beregnes den oprindelige og modificerede formel som følger:

$$\mathrm{S}_1 = \frac{\bar{d}}{\sqrt{\frac{2\pi \hat{f}_d(0)}{T}}} \tag{6.1}$$

$$\mathrm{S}_1^* = \left[\frac{n+1-2h+n^{-1}h(h-1)}{n}\right]^{1/2} S_1 \tag{6.2}$$

Udregningen af p-værdien i den modificerede test baserer sig på students T distributionen med $n-1$ frihedsgrader. Til beregningen af den modificerede Diebold & Mariano test og p-værdien benyttes "forecast" pakken i programmeringssproget R (Hyndman & Khandakar, 2008).

6.1 Generel model resultater

I tabel 6.1 fremgår den samlede forecast præcision for den konstruerede machine learning model samt analytikere generelt. Forecast præcisionen er jævnfør tabellen opgjort på baggrund af performance målene MAE, RMSE, MASE samt skill-score, og er endvidere oplyst for de respektive forecast horisonter på et, to og fire kvartaler. Resultaterne præsenteret i tabellen baserer sig endvidere på testsættet i machine learning modellen, bestående af observationer fra perioden 2018 – 2020 som belyst i afsnit 5.2.

Tabel 6.1: Generel forecast performance

	$h=1$		$h=2$		$h=4$	
	Analytiker	Model	Analytiker	Model	Analytiker	Model
MAE	0,53	0,46	0,62	0,48*	0,84	0,56**
RMSE	6,69	1,02	7,60	1,05	8,55	1,17
$\mathrm{SS}_{\mathrm{MAE}}$	0,32	0,41	0,25	0,43	0,26	0,51

Note: Tabellen angiver modellen og analytikeres forecast performance opgjort ved en forecast horisont på et, to og fire kvartaler. **, * indikerer hvorvidt modellens MAE er statistisk signifikant lavere end analytikeres på et 1% og 5% niveau.

6.1.1 Forecast horisont på 1 kvartal

Jævnfør tabel 6.1 fremgår det at den konstruerede machine learning model opnår en forecast præcision på 0,4582 opgjort ved performance målet MAE, ved en forecast horisont på et kvartal. Dette betyder, at modellens forecast i gennemsnit afviger 0,4582 i absolutte værdier fra de sande EPS værdier. Modellen formår dermed at outperforme analytikere med 13,42%, da analytikere tilsvarende

opnår en forecast præcision på 0,5293. På baggrund af performancemålet RMSE fremgår der en markant større forskel i forecast præcisionen mellem modellen og analytikere, end identificeret ved MAE. Modellen opnår en RMSE på 1,0247 mens analytikere har en RMSE på 6,6861. Da performance målet RMSE som belyst i afsnit 4.3.1 er særlig følsom overfor outliers, indikerer resultaterne alt andet lige, at forekommer en højere grad af outliers, i form af forecast fejl, i analytiker forecast end i modellen. Skill-scoren opgjort for analytikere og modellen reflekterer endvidere hvordan den gennemsnitlige forecast præcision er sammenlignet med et random walk forecast. Det fremgår på baggrund af tabellen, at modellen, ved en forecast horisont på et kvartal, har en gennemsnitlig forecast præcision der er 41,09% bedre end et naivt forecast. Det fremgår tillige, at analytikeres forecast er 31,95% bedre end et naivt forecast. På baggrund af performancemålene fremgår det, at den konstruerede machine learning model formår at outperforme analytikere ved en forecast horisont på et kvartal. På baggrund af en statistisk test af resultaterne kan det dog konstateres at modellens outperformance af analytikere ikke er statistisk signifikant, hvorfor det ikke kan udelukkes at modellens outperformance er tilfældig.

6.1.2 Forecast horisont på 2 kvartaler

På baggrund af resultaterne fremstillet i tabel 6.1, kan det konstateres at modellens outperformance af analytikere er mere markant ved en forlængelse af forecast horisonten fra et til to kvartaler. Endvidere formår modellen at outperforme analytikere med 23,50% baseret på MAE ved en forecast horisont på to kvartaler. Niveauet af RMSE blandt analytikere vidner om at der forekommer en højere grad af outliers i forecast fejl for analytikere sammenlignet med modellen, hvilket ligeledes tilfældet ved en horisont på et kvartal. Det fremgår endvidere at modellens skill-score forbedres marginalt sammenlignet med en forecast horisont på et kvartal, hvilket indikerer at modellens forecast præcision forbedres i forhold til et naivt forecast ved forlængelse af forecast horisonten. Modsat fremgår det blandt analytikere, at forecast præcisionen relativt til et naivt forecast forringes ved forlængelse af forecast horisonten, da skill-scoren falder fra 0,3195 til 0,2491 ved en forecast horisont på to kvartaler.

Det kan på baggrund af resultaterne ved en forecast horisont på to kvartaler konstateres at modellen opnår en markant bedre forecast præcision end analytikere baseret på de fremstillede performance mål. Endvidere kan det på baggrund af en statisk test af resultaterne konstateres, at modellens outperformance af analytikere er statistisk signifikant, hvorfor vi kan konkludere at modellens outperformance ikke er tilfældig. Modellens outperformance af analytikere bliver dermed statiske signifikant ved forlængelse af forecast horisonten fra et kvartal til to kvartaler.

6.1.3 Forecast horisont på 4 kvartaler

Ved en forecast horisont på fire kvartaler fremgår det, at modellens outperformance af analytikere stiger yderligere sammenlignet med forecast horisonten på to kvartaler. Modellens og analytikere opnår en forecast præcision opgjort ved MAE på henholdsvis 0,5573 og 0,8366, svarende til at modellen outperformer analytikere med 33,39%. Performance målet RMSE indikerer, ligesom det var tilfældet ved en forecast horisont på henholdsvis et og to kvartaler, at der forekommer en højere grad af outliers i forecast fejl for analytikere sammenlignet med modellen. På baggrund af modellens skill-score kan det konstateres at modellen outperformer et naivt forecast med 50,64%, hvilket er

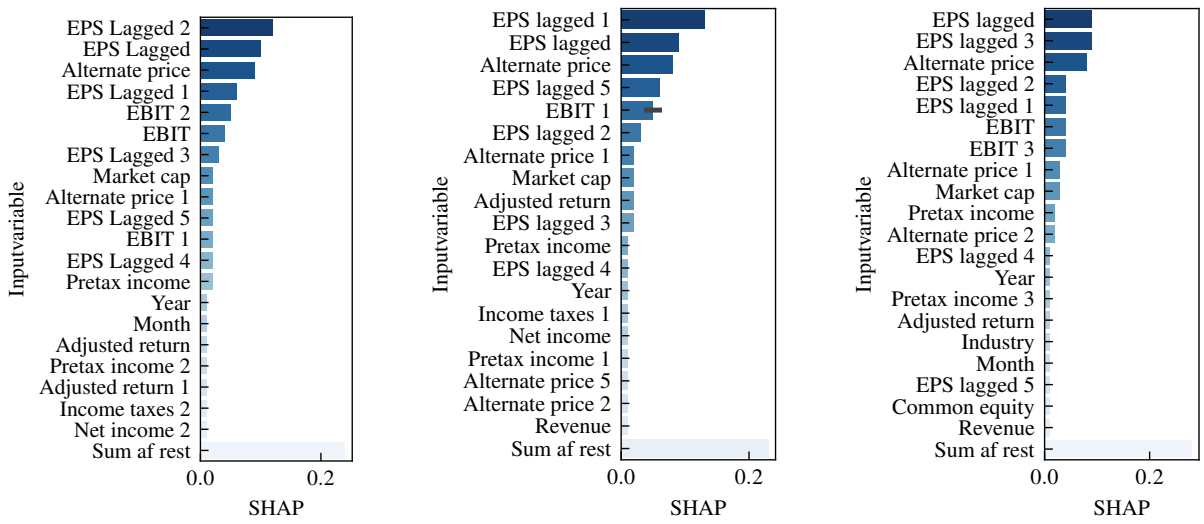
modellens største outperformance af et naivt forecast blandt de tre fremstillede forecast horisonter. Tilsvarende opnår analytikere en marginal stigning i skill-scoren sammenlignet med en forecast horisont på et kvartal, der på to kvartalers forecast horisont var 0,2591. Årsagen til forbedringen i forecast performance for både analytikere og modellen målt ved skill-score, kan alt andet lige tænkes at skyldes, at et naivt forecast forringes ved forlængelse af forecast horisonten, da virksomheder indtjening alt andet lige stiger over tid. På baggrund af resultaterne ved en forecast horisont på fire kvartaler, kan det med udgangspunkt i de anvendte performancemål konstateres at den konstruerede machine learning model, opnår den største outperformance af analytikere ved en forecast horisont på fire kvartaler. Endvidere afslører en statisk test af resultaterne, at modellens outperformance af analytikere var statistisk signifikant, med en p-værdi på 0,0035. De ekstreme outliers identificeret blandt analytikere, der kom til udtryk i performance målet RMSE, vil undersøges nærmere i det efterfølgende afsnit 6.2 med henblik på at klarlægge årsagen til disse.

6.1.4 Feature importance

I følgende afsnit vil forklaringsgraden af de respektive inputvariable i den konstruerede model evalueres, med henblik på at øge transparensen af modellen og dermed minimere black-box problemstillingen. I afsnittet vil inputvariablenes rangeres på baggrund af forklaringsgraden og det vil på baggrund heraf klarlægges, hvilke inputvariable der er mest udslagsgivende i modellens EPS forecast.

Machine learning algoritmer anses ofte for at producere black-box modeller, grundet den manglende forklaring af forecast. På baggrund heraf har forskere udviklet forskellige modelfortolkningsværktøjer, der gør det muligt at kvantificere samt visualisere forklaringsgraden af inputvariable i machine learning modeller (Casalicchio et al., 2019). For at kvantificere forklaringsgraden af modellens inputvariable i afhandlingen, benyttes SHAP values. SHAP values baserer sig på en spilteoretisk tilgang og har de seneste år vundet indpas hos teoretikere såvel som praktikere grundet sin anvendelighed på beslutningstræmodeller. Til beregning af Shap values i afhandlingen benyttes SHAP udviklet af Lundberg og Lee (2017) og den dertilhørende implementering i Python. I figur 6.1 fremgår forklaringsgraden af de respektive inputvariable opgjort ved en forecast horisont på henholdsvis et, to og fire kvartaler. Figuren illustrerer den gennemsnitlige absolutte værdi, de 20 inputvariable med størst forklaringsgrad bidrager med i det endelige forecast.

Figur 6.1: Feature importance



(a) $h = 1$ (b) $h = 2$ (c) $h = 4$

Note: Figuren illustrerer feature importance i modellen fordelt på forecast horisonter på et, to og fire kvartaler.

Det kan på baggrund af figur 6.1 konstateres, at inputvariablene med størst forklaringskraft overordnet set er meget ens på tværs af de tre forecast horisonter. Det fremgår endvidere at laggede EPS variable udgør de inputvariable med højest forklaringsgrad på tværs af alle forecast horisonter. Det kan hertil konstateres, at der er forskel på hvilket EPS tidslag, der har den største forklaringsgrad ved de forskellige forecast horisonter. Det fremgår blandt andet, at inputvariablen *EPS lagged 2* har den største forklaringsgrad ved en forecast horisont på et kvartal, og tilsvarende udgør *EPS lagged 1* og *EPS lagged* de inputvariable med størst forklaringsgrad ved en forecast horisont på henholdsvis to og fire kvartaler. På baggrund heraf kan det konstateres, at det EPS tidslag med størst forklaringskraft umiddelbart bestemmes ved, at den historiske EPS værdi forelægger præcist et år forud for tidspunktet der forecastes for. Dette gør sig blandt andet gældende ved en forecast horisont på henholdsvis 1 og 2 kvartaler, mens det ved en forecast horisont på fire kvartaler udgør den inputvariabel med næsthøjest forklaringskraft. Forklaringen på dette kan tænktes at skyldes, at visse selskabers indtjening er sæsonbetonet (e.g retail og manufacturing), hvortil realiseret EPS samme kvartal blot året før tidspunktet der forecastes, er indikativ for niveauet af EPS i det givne kvartal (Geremew & Gourio, 2018) Samme tendens gør sig gældende ved regnskabsvariablen *EBIT*, der ligeledes har en høj forklaringsgrad på tværs af de forskellige forecast horisonter.

Udover EPS lagged og EBIT fremgår det ydermere at aktiekursen, i form af inputvariablen *Alternate price*, har en høj forklaringsgrad ved alle forecast horisonter. Disse resultater understøtter ydermere argumenter fremstillet i litteraturen, der konkluderede at ændringer i aktiekurs er forklarende for selskabers indtjening (Ashton & Wang, 2013). Den væsentligste forskel i inputvariablenes forklaringsgrad på tværs af de forskellige forecast horisonter består i, at den samlede forklaringskraft udgøres af flere inputvariable ved en forecast horisont på fire kvartaler, sammenlignet med forecast horisonten på henholdsvis et og to kvartaler. Dette kommer endvidere til udtryk ved at forklaringsgraden for *Sum af rest* kategorien er større ved en forecast horisont på fire kvartaler, sammenlignet med de to andre forecast horisonter. Resultaterne heraf indikerer alt andet lige, at der

er flere faktorer, der er afgørende for forecastet ved længere forecast horisonter end ved korte forecast horisonter.

6.2 Industri

I følgende afsnit undersøges machine learning modellens anvendelighed på tværs af industrier med henblik på at besvare arbejdsspørgsmål 2. Yderligere vil det i afsnittet vurderes, hvorvidt det er mest hensigtsmæssigt at implementere industrispecifikke modeller fremfor en generel model. Afslutningsvis sammenlignes modellens performance med analytikeres for de respektive industrier.

6.2.1 Industrispecifikke machine learning modeller

Analytikere specialiserer sig hovedsageligt inden for enkelte industrier, for at opnå industrispecifik viden, der kan benyttes i indtjeningsestimater, aktiekursestimater samt i forbindelse med udarbejdelse af analyserapporter. I studiet af Brown et al (2014) fremgik det, at analytikere vurderede deres industri viden, som det vigtigste input til at estimere selskabers fremtidige indtjening. Ydermere fremgik det, at analytikeres industrividen udgjorde det væsentligste parameter til at forklare den enkelte analytikers lønkompensation, hvilket understreger vigtigheden af analytikeres industrispecifikke viden. Givet vigtigheden af industrispecifik viden tidligere identificeret i litteraturen, er det endvidere relevant at undersøge hvorvidt den konstruerede machine learning model forecast præcision forbedres, såfremt modellen trænes på hver enkelt industri fremfor hele datasættet. Træningen af modellen baseret på de individuelle industrier vil alt andet lige give modellen bedre betingelser for at identificere industrispecifikke mønstre og tendenser sammenlignet med en generel model.

I litteraturen er det identificeret at regnskabsposters forklaringsgrad for selskabers indtjening kan variere markant på tværs af forskellige industrier. Endvidere kunne Seng & Hancock (2012) i sit studie konstatere, at varelageret var statistisk signifikant til at forklare indtjeningen i industrierne manfacturing og primary products, mens varelagerets forklaringsgrad ikke var statistisk signifikant i industrien retail. Ydermere fremgik det ligeledes at anlægsinvesteringer og tilgodehavende, havde en forklaringsgrad der varierede væsentligt på tværs af forskellige industrier. Resultaterne fremstillet i litteraturen understreger dermed fundamentale forskelle i regnskabsvariables betydning for indtjening i forskellige industrier, hvorfor afhandlingen vil undersøge, hvorvidt en machine learning model trænet på hver enkelt industri performer bedre end en generel model.

Generel model vs. industrispecifikke modeller

Opdelingen af industrier foretages på baggrund af Fama & Frenchs definitioner, som beskrevet i afsnit . Det er hertil fundet hensigtsmæssigt at ekskludere industrierne Utilities og Other, da Utilities industrien kun består af 507 observationer, hvilket ikke er tilstrækkeligt givet den påkrævede opdeling i trænings- og testdata. Ydermere er Other ekskluderet fra afsnittet, da denne ikke repræsenterer én specifik industri, men derimod udgør en kombination af forskellige industrier. I

tabel 6.2 fremgår forecast præcisionen opgjort ved en generel model samt en industrispecifik model, ved en forecast horisont på henholdsvis et, to og fire kvartaler.

Tabel 6.2: Industrispecifikke modeller Vs. generel model

	$h = 1$		$h = 2$		$h = 4$	
Industri	Forskel (%)	Input	Forskel (%)	Input	Forskel (%)	Input
Healthcare	2,35*	8	0,92	7	0,24	8
Manufacturing	1,20	8	1,21	9	0,00	7
Business Equipment	2,46**	9	1,98**	9	0,03	9
Wholesale/Retail	3,04*	8	0,62	8	-0,21	7
Finance	5,42**	9	3,65*	7	0,09	6
Chemicals & A.P	3,92*	8	7,17**	8	4,05	7
Consumer N.D	4,86**	8	3,14	8	2,64	8
Energy	-0,81	6	-0,61	7	-7,00	7
Telecom	5,70*	8	5,91*	8	-0,03	7
Consumer durables	6,80**	7	-0,92	8	-2,24	8

Note: Tabellen angiver forskellen i MAE mellem de industrispecifikke modeller og den generelle model opgjort i %, opgjort som (Industrimodel / generel model -1). Kolonnen *Input* angiver antallet af top 10 inputvariable, hver industrimodel deler med den generelle model. **, * indikerer hvorvidt den generelle models MAE er statistisk signifikant lavere end de industrispecifikke modeller på et 1% og 5% niveau.

Ved en forecast horisont på et kvartal fremgår det, på baggrund af tabellen, at den generelle model opnår en højere forecast præcision end den industrispecifikke model i alle industrier, med undtagelse af industrien Energy. Den generelle models outperformance af den industrispecifikke model er mest markant i industrierne Consumer durables, Telecom og Finance, hvor den generelle model opnår en outperformance på henholdsvis 6,80%, 5,70% og 5,42%. Endvidere kan det på baggrund af en statistisk test af resultaterne konkluderes, at den generelle models outperformance ved en forecast horisont på et kvartal er statistisk signifikant i alle industrier bortset fra i Energy og Manufacturing industrierne. En mulig årsag til den generelle models outperformance i størstedelen af industrierne, kan tænkes at være, at opdelingen af datasættet i industrier medfører, at mængden af træningsdata i den industrispecifikke model ikke er tilstrækkelig. Dette betyder dermed, at den industrispecifikke machine learning model ikke identificerer mønstre og tendenser i datagrundlaget i samme omfang som den generelle model.

Ved en horisont på henholdsvis to og fire kvartaler er konklusionerne mere uklare. Den generelle model formår at outperforme den industrispecifikke model i otte ud af de ti industrier, men viser sig dog kun statistisk signifikant i fem af disse industrier, ved en forecast horisont på to kvartaler. Sammen tendens gør sig gældende ved en forecast horisont på fire kvartaler, hvor det fremgår at den generelle model formår at outperforme industrimodellen i seks ud af de ti industrier. Den generelle models outperformance viser sig dog ikke statistisk signifikant i disse industrier, hvorfor resultaterne adskiller sig markant fra resultaterne ved en forecast horisont på et kvartal, hvor den generelle models outperformance var statistisk signifikant i otte ud af de ti industrier.

For undersøge forskellen mellem en generel model og en industrispecifik model yderligere, vil forskelle i inputvariablenes forklaringskraft i de forskellige modeller analyseres. På baggrund af de

respektive inputvariables forklaringskraft, som beskrevet i afsnit 6.1.4, vil det undersøges, hvor mange af de ti inputvariable med størst forklaringsgrad, industrimodellen deler med den generelle model, ved en forecast horisont på henholdsvis et, to og fire kvartaler. I tabel 6.2 fremgår antallet af de top ti inputvariable den industrispecifikke model deler med den generelle model fordelt på de respektive industrier. På baggrund af tabellen kan det konstateres, at de industrispecifikke modeller overordnet set deler mellem syv til 9 inputvariable med den generelle model. Det fremgår endvidere, at de industrispecifikke modeller deler færre inputvariable med den generelle model ved en forecast horisont på fire kvartaler sammenlignet med forecast horisonterne på et og to kvartaler. Dette kan tænkes at forklares af, at modellen har nemmere ved at identificere industrispecifikke mønstre på en længere forecast horisont, hvilket ligeledes fremgik Becker et al. (2004) resultater. Becker et al. (2004) konstaterede tilsvarende at de største industri forskelle i forecast fejl forekom på længere forecast horisonter.

Det blev identificeret i tabel 6.2 at Energy industrien var den eneste industri, hvor den industri specifikke model konsekvent outperformede den generelle model på tværs af de tre horisonter. Det fremgår endvidere at særligt Energy industrien, deler få inputvariable med den generelle model, med henholdsvis seks, syv og syv fælles inputvariable. Dette indikerer alt andet lige, at særligt Energy industrien har industrispecifikke mønstre og tendenser, som en generel model ikke opfanger i samme omfang som en industrispecifik model.Ydermere er resultaterne i Finance industrien særligt bemærkelsesværdige, da der fremgår en markant forskel i antallet af inputvariable, der deles med den generelle model, på tværs af de respektive forecast horisonter. Ved en forecast horisont på henholdsvis et kvartal fremgår det, at industrimodellen deler ni inputvariable med den generelle model, mens at industrimodellen ved en forecast horisont på henholdsvis to og fire kvartaler kun deler syv og seks inputvariable. Dette indikerer at den industrispecifikke model i Finance kategorien i højere grad lægger vægt på andre inputvariable end den generelle model, ved forlængelse af forecast horisonten. Set i lyset af resultaterne fremstillet i tabel 6.2, er disse resultater særligt interessante, da det endvidere fremgår at den generelle models outperformance af den industrispecifikke model dermed falder i takt med at antallet af fælles inputvariable falder.

Overordnet set kan det konkluderes, at de industrispecifikke modeller hovedsageligt inddrager de samme inputvariable med størst forklaringskraft, som den generelle model. Endvidere vil en mærkbar forskel ved implementering af industrispecifikke modeller for så vidt kræve, at industrispecifikke inputvariable. Grundet afhandlingens fokus på at konstruere en machine learning model baseret på regnskabstal og aktiedata, er inddragelse af industrispecifikke inputvariable uden for afhandlingens scope.

6.2.2 Model vs. analytikere

Resultaterne fremstillet i afsnit 6.2.1 viste, at en generel machine learning model trænet på hele datasættet, opnåede en statistisk signifikant outperformance af industrispecifikke modeller i størstedelen af industrierne, ved en forecast horisont på henholdsvis et og to kvartaler. Ydermere outperformede den generelle model de industrispecifikke modeller ved en forecast horisont på fire kvartaler, hvor outperformance dog ikke viste sig at være statistisk signifikant. På baggrund heraf, vil den generelle models forecast præcision derfor sammenlignes med analytikere for hver af de 11[12] industrier, med henblik på, at identificere potentielle forskelle i modellen og analytikeres

[12]Utilities er udeladt, da denne industri kun indeholder 27 observationer fordelt på 4 virksomheder i testperioden.

performance på tværs af industrierne. I tabel 6.3 er henholdsvis modellens og analytikeres forecast præcision opgjort for hver industri, ved en forecast horisont på et kvartal.

Forecast horisont på 1 kvartal

Tabel 6.3: Forecast performance opgjort pr. industri, $h = 1$

	Analytiker			Model			
Industri	MAE	RMSE	SS_{MAE}	MAE	RMSE	SS_{MAE}	Dækning
Healthcare	0,74	14,85	-0,02	0,35	0,76	0,52	5
Manufacturing	0,53	1,14	0,42	0,52	1,07	0,44	6
Business Equipment	0,27	0,66	0,35	0,28	0,62	0,33	7
Wholesale/Retail	0,50	1,22	0,36	0,49	1,01	0,36	7
Finance	0,45	1,18	0,36	0,51	1,16	0,29	6
Chemicals & A.P	0,56	1,11	0,30	0,51*	1,07	0,35	8
Consumer N.D	0,55	1,31	0,34	0,55	1,25	0,34	7
Energy	1,12	2,24	0,52	1,06	1,94	0,55	13
Telecom	0,32	0,65	0,46	0,36	0,81	0,39	8
Consumer durables	0,58	1,20	0,26	0,56	1,13	0,29	5
Other	0,46	1,08	0,36	0,52	1,16	0,29	6

Note: Tabellen angiver modellen og analytikeres forecast performance opgjort ved en forecast horisont på et kvartal. Kolonnen *Dækning* angiver median analytikerdækningen i hver industri. **, * indikerer hvorvidt modellens MAE er statistisk signifikant lavere end analytikeres på et 1% og 5% niveau.

På baggrund af tabellen kan det konstateres at modellens forecast præcision er bedre end analytikeres i syv ud af de 11 industrier. En statistisk test af resultaterne afslører dog, at modellens outperformance af analytikere kun er statistisk signifikant i industrien Chemicals and Allied products. Endvidere formår analytikere at outperforme modellen i industrierne Business Equipment, Finance, Telecom og Other med henholdsvis 3,48%, 11,55%, 13,31% og 11,71%. Modellens outperformance er mest markant i Healthcare industrien, hvor modellens forecast præcision er 52,50% bedre end analytikeres. Ydermere fremgår det på baggrund af skill-scoren at modellen outperformer et naivt forecast med 51,50%, hvor analytikere tilsvarende opnår en forecast præcision der er 2,12% dårligere. I de resterende seks industrier, hvor modellen outperformer analytikere er modellens outperformance i intervallet 0,51% til 7,79%. På baggrund af performance målet RMSE kan det konstateres, at analytikeres forecast fejl i Healthcare industrien bærer præg af ekstreme outliers, da det fremgår at analytikere opnår en RMSE på 14,85 sammenlignet med 0,76 på modellen.

I tabel 6.3 fremgår median analytikerdækningen på tværs af de 11 industrier, hvortil det fremgår antallet af analytikere der dækker hver virksomhed, er i spændet fem til 13. Endvidere udgør Energy industrien, den industri med den største analytiker dækning, hvortil det er særligt bemærkelsesværdigt at modellen formår at outperforme, da en høj analytiker dækning alt andet lige bør afspejle et robust analytiker estimat. Ydermere kan det på baggrund af skill-scoren konstateres, at de industrier modellen har sværest ved at forecaste er henholdsvis Finance, Other og Consumer durables, hvor modellen kun formår at outperforme et naivt forecast med 28,52%, 28,80% og

28,87%. Modsat vidner skill-scoren om, at analytikeres forecast præcision er dårligst i industrierne Healthcare, Consumer durables og Chemicals and Allied products, hvor analytikere outperformer et naivt forecast med henholdsvis -2,12%, 25,90% og 29,83%. Med undtagelse at Finance, Telecom, Other og Healthcare industrierne er niveauet for skill-scoren for analytikere og modellen tæt på hinanden. Dette indikerer alt andet lige, at de industrier hvor modellen opnår en høj forecast præcision, sammenlignet med et naivt forecast, ligeledes er industrier, hvor analytikere opnår en høj forecast præcision. Omvendt indikerer dette også, at industrier hvor modellens forecast præcision er mindre god i forhold til et naivt forecast, generelt er industrier, der er vanskelige at forecaste, da analytikere ligeledes opnår en lavere forecast præcision i disse industrier.

Forecast horisont på 2 kvartaler

Tabel 6.4: Forecast performance opgjort pr. industri, $h = 2$

	Analytiker			Model			
Industri	MAE	RMSE	SS_{MAE}	MAE	RMSE	SS_{MAE}	Dækning
Healthcare	0,96	16,96	0,22	0,37	0,78	0,70	5
Manufacturing	0,59	1,26	0,32	0,53**	1,09	0,39	5
Business Equipment	0,31	0,86	0,26	0,29*	0,64	0,31	7
Wholesale/Retail	0,55	1,32	0,32	0,51	1,04	0,37	6
Finance	0,46	1,12	0,28	0,49	1,10	0,23	6
Chemicals & A.P	0,64	1,25	0,15	0,54**	1,15	0,29	8
Consumer N.D	0,53	1,12	0,32	0,53	1,09	0,32	7
Energy	1,34	3,09	0,24	1,14**	2,04	0,36	12
Telecom	0,35	0,70	0,21	0,40	0,97	0,09	7
Consumer durables	0,64	1,32	0,15	0,60**	1,21	0,21	5
Other	0,55	1,32	0,21	0,54	1,22	0,22	5

Note: Tabellen angiver modellen og analytikeres forecast performance opgjort ved en forecast horisont på to kvartaler. Kolonnen *Dækning* angiver median analytikerdækningen i hver industri. **, * indikerer hvorvidt modellens MAE er statistisk signifikant lavere end analytikeres på et 1% og 5% niveau.

Resultaterne i tabel 6.4 viser henholdsvis modellens og analytikeres performance opgjort for hver enkelt industri, ved en forecast horisont på to kvartaler. Resultaterne fremstillet i tabellen indikerer endvidere at modellens outperformance af analytikere er mere markant, set i forhold til en forecast horisont på et kvartal, hvilket er konsistent med resultaterne fremstillet i afsnit 6.1.

På baggrund af performancemålet MAE, kan det konstateres, at modellen outperformer analytikerne på ni ud af de 11 industrier. Endvidere formår modellen ikke at outperforme analytikere i industrierne Finance og Telecom, hvilket ligeledes var tilfældet ved en forecast horisont på et kvartal, hvor analytikere outperformer modellen med henholdsvis 6,93% og 16,07%. En statistisk test afslører endvidere, at modellens outperformance af analytikere er statistisk signifikant i fire ud af de 11 industrier, hvilket er en forbedring set i forhold til resultaterne ved en forecast horisont på et kvartal, hvor modellens outperformance kun var statistik signifikant for en industri. Ydermere fremgår det, at Energy industrien, som det var tilfældet ved en forecast horisont på et kvartal, har den højeste

analytikerdækning blandt industrierne, med en median dækning på 12 analytikere. Til trods for den høje analytikerdækning formår modellen at outperforme analytikere med 15,36%. På baggrund af RMSE kan det ydermere konstateres, at særligt Healthcare og Energy industrierne bærer præg af en høj grad af outliers i forecast fejlene. Det fremgår endvidere at graden af outliers er størst blandt analytikere i disse industrier, med en RMSE på 16,96 og 3,09.

Skill-scoren vidner om, at industrierne Consumer durables og Other er blandt de industrier, hvor både modellen og analytikere opnår den laveste outperformance af et naivt forecast. Endvidere formår modellen at outperforme det naive forecast med 21,2% og 22,18%, mens analytikere tilsvarende opnår en outperformance på henholdsvis 15,45% og 21,2%. Resultaterne kan indikere, at særligt disse industrier er svære at estimere, eller at industrierne bærer præg af en relativt stabil indtjening, hvorfor et naivt forecast er relativt præcist. Med undtagelse af Healthcare og Telecom, er niveauet af skill-scoren forholdsvis identisk på både analytikerne og modellen, hvilket ligeledes var tilfældet ved en forecast horisont på et kvartal.

Forecast horisont på 4 kvartaler

Tabel 6.5: Forecast performance opgjort pr. industri, $h = 4$

	Analytiker			Model			
Industri	MAE	RMSE	SS_{MAE}	MAE	RMSE	SS_{MAE}	Dækning
Healthcare	1,60	19,59	0,42	0,42*	0,87	0,85	3
Manufacturing	0,72	1,40	0,13	0,58**	1,14	0,30	4
Business Equipment	0,39	1,00	0,15	0,35**	0,72	0,24	6
Wholesale/Retail	0,63	1,28	0,21	0,58**	1,08	0,27	4
Finance	0,46	1,00	0,15	0,50	0,98	0,07	4
Chemicals & A.P	0,79	1,46	-0,02	0,62**	1,32	0,19	5
Consumer N.D	0,63	1,22	0,19	0,59*	1,15	0,23	5
Energy	1,75	2,89	0,03	1,44**	2,45	0,20	9
Telecom	0,38	0,72	0,47	0,38	0,72	0,47	5
Consumer durables	0,82	1,66	0,04	0,73**	1,47	0,15	3,5
Other	0,67	1,44	0,11	0,64**	1,32	0,14	4

Note: Tabellen angiver modellen og analytikeres forecast performance opgjort ved en forecast horisont på fire kvartaler. Kolonnen *Dækning* angiver median analytikerdækningen i hver industri. **, * indikerer hvorvidt modellens MAE er statistisk signifikant lavere end analytikeres på et 1% og 5% niveau.

Identisk med en forecast horisont på to kvartaler formår modellen at outperforme analytikere i ni ud af 11 industrier. Endvidere formår modellen ikke at outperforme analytikere i Finance og Telecom industrierne, hvilket er de industrier, hvor analytikere er modellen overlegen på tværs af alle forecast horisonter. Modellens outperformance af analytikere er statistisk signifikant i alle ni industrier, og opnår endvidere en p-værdi på under 0,01 i syv industrierne. Resultaterne vidner dermed om en konsekvent outperformance af analytikere på denne forecast horisont. Konsistent med resultaterne ved en forecast horisont på et og to kvartaler, opnår modellen en markant outperformance af analytikere i Healthcare industrien, hvor modellen formår at outperforme analytikere med 73,47%.

Udover Healthcare industrien opnår modellen ligeledes en høj outperformance af analytikere i industrierne Chemicals and Allied products og Manufacturing, hvilket ligeledes gjorde sig gældende på de andre forecast horisonter, med en outperformance på henholdsvis 20,94% og 19,44%. Som tilfældet ved forecast horisonterne på henholdsvis et og to kvartaler vidner RMSE i tabellen, om ekstreme outliers i analytikeres forecast fejl i Healthcare industrien.

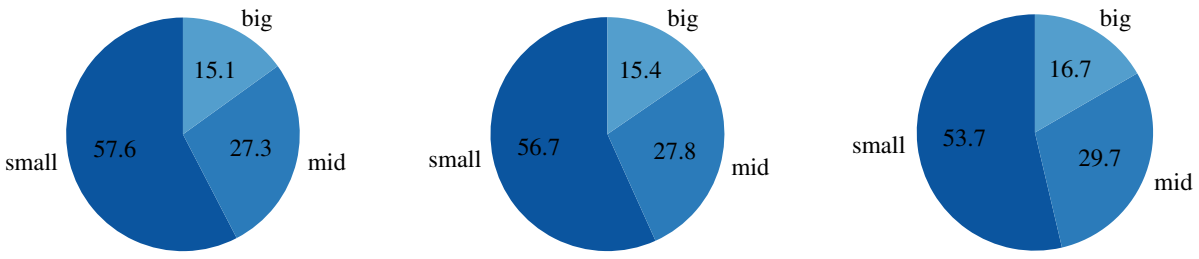
På baggrund af skill-scoren kan det endvidere konstateres, at modellen er dårligst til at forecaste industrierne Finance, Consumer durables og Other, mens analytikere tilsvarende performer dårligst i Chemicals and Allied products, Energy og Consumer durables industrien. Særligt bemærkelsesværdig er forskellen mellem modellen og analytikeres skill-score i Chemicals and Allied products industrien, hvor modellen outperformer et naivt forecast med 19,41% mens analytikeres forecast præcision er 1,93% dårligere end et naivt forecast. Ydermere vidner resultaterne om væsentlige forskelle i skill-scoren mellem analytikere og modellen på tværs af industrier, hvilket ikke forekom i samme omfang ved en forecast horisont på et og to kvartaler. De markante forskelle i skill-scoren skyldes modellens dominerende forecast præcision ved længere forecast horisonter, hvilket ligeledes kunne konstateres i afsnit 6.1.

6.3 Market cap

I det følgende afsnit vil det klarlægges, hvilken indflydelse virksomhedsstørrelsen har på henholdsvis analytiker og modellens forecast præcision i afhandlingen, med henblik på at besvare arbejdsspørgsmål 3 i den fremstillede problemformulering.

I litteraturen har tidligere studier kunne konstatere en sammenhæng mellem virksomhedsstørrelsen opgjort ved markedsværdien, og analytikeres forecast præcision. Brown et al. (1987) resultater viste en positiv sammenhæng mellem virksomhedsstørrelsen og analytikeres outperformance af statiske modeller. Ydermere konstaterede Ball & Ghysels (2017), at deres MIDAS model opnåede en højere forecast præcision sammenlignet med analytikere, når der var tale om mindre virksomheder. For at klarlægge, hvorvidt der er forskel i analytikere og modellens forecast præcision på baggrund af virksomhedsstørrelser, er det fundet hensigtsmæssigt at opdele observationerne i datagrundlaget i de tre kategorier; *small-cap, mid-cap* og *large-cap*. Inddelingen i de tre kategorier baserer sig endvidere på grænseværdierne fremstillet af CRSP databasen (CRSP, 2021). I figur 6.2 illustreres fordelingen af observationer i de tre kategorier i datagrundlaget, baseret på markedsværdien af de enkelte selskaber.

Figur 6.2: Fordeling af observation på virksomhedsstørrelse



(a) $h = 1$ (b) $h = 2$ (c) $h = 4$

Note: Figuren angiver fordelingen af observationer på virksomhedsstørelse, ved en forecast horisont på et, to og fire kvartaler. De forskellige Kategorier er opdelt på baggrund af følgende grænseværdier *small-cap: <2*, *mid-cap lav: 2-10*, *large-cap: 10+*.

Det fremgår af figuren, at fordelingen af observationer i de tre forskellige kategorier af virksomhedsstørrelser er mere eller mindre identisk på tværs af de tre forecast horisont. Endvidere udgør small-cap kategorien den største kategori med omtrent 55% af alle observationer, efterfulgt at mid-cap med 28% og large-cap med 16%.

I tabel 6.6 fremgår forecast præcisionen ved en forecast horisont på et kvartal, opdelt på baggrund af de tre kategorier small-cap, mid-cap og large-cap. I tabellen er forecast præcisionen opgjort for både modellen samt analytikere, og ydermere fremgår medianen af antallet af analytikere, der dækker selskaberne i hver kategori.

Tabel 6.6: Forecast performance opgjort på market cap, $h = 1$

	Analytiker		Model		
	MAE	RMSE	MAE	RMSE	Dækning
Small cap	0,57	8,76	0,43	1,01	4
Mid cap	0,46	1,10	0,49	1,03	9
Large cap	0,51	1,13	0,53	1,07	16

Note: Tabellen angiver modellen og analytikeres forecast performance opgjort ved en forecast horisont på et kvartal. Kolonnen *Dækning* angiver median analytikerdækningen fordelt på hver market cap kategori. **, * indikerer hvorvidt modellens MAE er statistisk signifikant lavere end analytikeres på et 1% og 5% niveau.

Det kan på baggrund af tabel 6.6 konstateres, at modellen, ved en forecast horisont på et kvartal, opnår en højere forecast præcision end analytikere i én ud af de tre kategorier. Det fremgår endvidere, at modellen opnår en MAE på 0,4259 i kategorien small-cap, hvortil analytikere tilsvarende opnår en MAE på 0,5693. Dette betyder dermed, at modellen formår at outperforme analytikere med 25,18% i denne kategori. Modsat resultaterne i small-cap kategorien kan det konstateres, at modellen ikke formår at outperforme analytikere i kategorierne mid-cap og large-cap, hvor modellens forecast

præcision er henholdsvis 6,87% og 3,91% lavere end analytikernes. På baggrund af performancemålet RMSE fremgår det, at der forekommer en højere grad af outliers i analytikeres forecast fejl sammenlignet med modellen, ligesom tidligere konstateret. Det hertil relevant at bemærke, at graden af outliers i analytikeres forecast fejl er særlig markant i kategorien small-cap, hvortil der fremgår en RMSE på 8,757. Som belyst i afsnit 6.2, fremgik det, at særligt Healtcare industrien oplevede en høj RMSE. For at identificere hvorvidt outliers forekommer generelt i small-cap kategorien, eller om det isoleret set skyldes selskaber fra Healthcare industrien, er der foretaget en genberegning af RMSE ved ekskludering af Healthcare selskaber i bilag C.1. Resultaterne i bilaget indikerer endvidere, at den ekstreme RMSE identificeret i small-cap kategorien isoleret stammer fra Healtcare selskaber. Ved ekskludering af Healtcare selskaber fremgår en RMSE på 1,15 blandt analytikere, hvilket er omtrent samme niveau som RMSE i henholdsvis mid- og large-cap kategorien. På baggrund af tabel 6.6 kan det ydermere konstateres, at der er forskel i analytikerdækningen på tværs af de forskellige kategorier. Det fremgår endvidere, at small-cap kategorien har den laveste analytiker dækning, med en median på fire analytikere, mens henholdsvis mid-cap og large-cap kategorierne har en tilsvarende analytikerdækning på henholdsvis ni og 16 analytikere. Til trods for modellens markante outperformance af analytikere i small-cap kategorien, er modellens outperformance ikke statistisk signifikant, hvorfor det ikke kan udelukkes at der er tale om tilfældigheder.

Tabel 6.7: Forecast performance opgjort på market cap, $h = 2$

	Analytiker		Model		
	MAE	RMSE	MAE	RMSE	Dækning
Small cap	0,67	10,01	0,43*	1,01	4
Mid cap	0,54	1,48	0,51	1,08	8
Large cap	0,61	1,42	0,57*	1,17	14

Note: Tabellen angiver modellen og analytikeres forecast performance opgjort ved en forecast horisont på to kvartaler. Kolonnen *Dækning* angiver median analytikerdækningen fordelt på hver market cap kategori. **, * indikerer hvorvidt modellens MAE er statistisk signifikant lavere end analytikeres på et 1% og 5% niveau.

Modsat resultaterne ved en forecast horisont på et kvartal, fremgår det jævnfør tabel 6.7 at modellen outperformer analytikere i alle kategorier, ved en forecast horisont på to kvartaler. Det fremgår endvidere, at modellens outperformance af analytikere er mest markant i small-cap kategorien, hvor modellens outperformance er 35,20%. Ydermere formår modellen at outperforme analytikere i kategorierne mid-cap og large-cap med henholdsvis 4,93 og 6,41%.

Som identificeret ved en forecast horisont på et kvartal, kan den ekstreme værdi af RMSE i small-cap kategorien isoleres til Healtcare industrien jævnfør bilag C.1. På baggrund af tabellen kan det konstateres, at analytikerdækningen er uændret i small-cap kategorien, mens kategorierne mid-cap og large-cap oplever fald på henholdsvis et og to analytikere i analytikerdækningen, set i forhold til en forecast horisont på et kvartal. På baggrund af en statiske test af resultaterne, kan det konkluderes at modellens outperformance af analytikere er statistisk signifikant i kategorierne small-cap og large-cap selskaber. Modellens outperformance af analytikere viser sig dermed at være insignifkant i mid-cap kategorien, men vil med en p-værdi på 0,0634 kategoriseres som grænse insignifikant.

Tabel 6.8: Forecast performance opgjort på market cap, $h = 4$

	Analytiker		Model		
	MAE	RMSE	MAE	RMSE	Dækning
Small cap	0,94	11,58	0,47**	1,06	3
Mid cap	0,68	1,37	0,62**	1,22	6
Large cap	0,80	1,72	0,71**	1,39	9

Note: Tabellen angiver modellen og analytikeres forecast performance opgjort ved en forecast horisont på fire kvartaler. Kolonnen *Dækning* angiver median analytikerdækningen fordelt på hver market cap kategori. **, * indikerer hvorvidt modellens MAE er statistisk signifikant lavere end analytikeres på et 1% og 5% niveau.

Ved en forecast horisont på fire kvartaler fremgår det, at modellen opnår en bedre forecast præcision end analytikere på tværs af alle de fremstillede kategorier. På baggrund af tabel 6.8, kan det endvidere konstateres at modellens outperformance af analytikere er mest markant i small-cap kategorien med en outperformance på 49,34%, hvilket er konsistent med resultaterne fremstillet ved de andre forecast horisonter. Den ekstreme RMSE, der forekommer i small-cap kategorien kan, som det var tilfældet ved tidligere forecast horisonter, isoleres til Healthcare industrien. Det fremgår endvidere, at analytikerdækningen falder i alle kategorier ved forlængelse af forecast horisonten fra to til fire kvartaler, hvortil det største fald forekommer i large-cap kategorien med et fald på fem analytikere. På baggrund af en statisk test, kan det konkluderes, at modellens outperformance af analytikere er statistisk signifikant i alle kategorier.

Resultaterne fremstillet for de respektive kategorier på tværs af de forskellige forecast horisonter vidner endvidere om, at den konstruerede machine learning model opnår en særdeles god forecast præcision sammenlignet med analytikere i small-cap kategorien. Dette kommer til udtryk ved, at modellen formår at outperforme analytikere på alle forecast horisonter, der endvidere er statistisk signifikant ved en forecast horisont på henholdsvis to og fire kvartaler. Der kan være adskillelige årsager til, at modellens outperformance af analytikere er særlig fremtrædende i small-cap kategorien. Tidligere studier identificerede en positiv sammenhæng mellem virksomhedsstørrelsen og offentliggørelsen af virksomhedsinformation, hvortil det kunne konstateres, at kvaliteten og detaljeringsgraden af offentliggjort information er bedre blandt større virksomheder (Leung & Srinidhi, 2006; M. Madhani, 2016). Dette kan endvidere være en af årsager til den markante outperformance af analytikere i small-cap kategorien, da der ifølge Holland (1998) fremgik en højere forecast præcision i analytikeres estimater, såfremt der forelå detaljeret information i forlængelse af

finansielle opgørelser, i form af beskrivende og forklarende noter. Ydermere indikerede resultater fremstillet af Lang & Lundholm (1996) en positiv relation mellem analytikerdækning og analytikeres forecast præcision, hvilket ligeledes kan være en af årsagerne til modellens markante outperformance af analytikere i small-cap kategorien.

6.4 Analytikerdækning og spredning i estimater

I følgende afsnit vil analytikerdækning samt spredningen i analytiker estimaters indflydelse på modellens outperformance af analytikere analyseres, med henblik på at besvare arbejdsspørgsmål 3 i afhandlingens problemstilling.

6.4.1 Analytikerdækning

På baggrund af resultaterne i afsnit 6.3 fremgik det, at den konstruerede machine learning models outperformance af analytikere var mest markant i small-cap kategorien, hvor der endvidere kunne konstateres den laveste analytikerdækning. Set i lyset af resultaterne fremstillet i litteraturen, hvor Lang & Lundholm (1996) konstaterede en positiv sammenhæng mellem antallet af analytikere, der dækker selskaberne og analytikeres forecast præcision, er det hertil relevant at undersøge, i hvilket omfang analytikerdækningen påvirker modellens outperformance af analytikere. Vi ønsker endvidere at analysere, hvorvidt modellen kun er i stand til at outperforme analytikere såfremt der er en lav analytikerdækning, eller om dette ligeledes gør sig gældende ved en høj analytikerdækning.

For at vurdere analytikerdækningens indflydelse på modellens outperformance af analytikere inddeles analytikerdækningen i de fire overordnede kategorier; *lav*, *mellem lav, mellem høj, høj,* på baggrund af antallet af analytikere der dækker selskabet. I tabel 6.9 fremgår modellens outperformance af analytikere opgjort for hver kategori, samt på forecast horisonter på henholdsvis et, to og fire kvartaler.

Tabel 6.9: Modellens outperformance af analytikere opgjort på analytikerdækning

	$h = 1$	$h = 2$	$h = 4$
	Outperformance %	Outperformance %	Outperformance %
Lav	57,40	58,76	58,20*
Mellem lav	-0,31	14,38*	31,42*
Mellem høj	-2,57	5,85*	11,01**
Høj	-4,41**	7,40**	12,14**

Note: Tabellen angiver modellens procentuelle outperformance af analytikere i forskellige kategorier af analytikerdækning, opgjort på baggrund af MAE. Kategorierne er opdelt på baggrund af følgende grænseværdier *lav: 1-2*, *Mellem lav: 3-5*, *Mellem høj: 6-8* og *Høj: 9+*. Ved en forecast horisont på et kvartal indikerer **, * hvorvidt analytikeres outperformance af modellen er statstisk signifikant forskellig fra 0 på et 1% og 5% niveau. Ved en forecast horisont på to og fire kvartaler indikerer **, * hvorvidt modellens outperformance er statistisk signifikant på et 1% og 5% niveau.

Det kan på baggrund af resultaterne konstateres, at modellen kun formår at outperforme analytikere i én ud af de fire fremstillede kategorier, ved en forecast horisont på et kvartal. Det fremgår at modellen outperformer analytikere med 57,4% i kategorien *lav*. Endvidere formår modellen ikke at outperforme analytikere i kategorierne *mellem lav, mellem høj* og *høj*. I disse kategorier fremgår det, at analytikeres forecast præcision er marginalt bedre end modellen, med en outperformance af modellen på henholdsvis 1,09%, 2,57% og 4,41%. På baggrund af en statisk test kan det konkluderes, at analytikeres og modellens performance kun er statistisk forskellig fra 0 i kategorien *høj*, hvorfor vi ikke kan afvise, at forskellen i performance mellem analytikere og modellen på *lav, mellem lav* og *mellem høj* er tilfældig. Ved en forecast horisont på henholdsvis to og fire kvartaler fremgår en markant ændring i analytikeres outperformance. Modellen formår endvidere, at outperforme analytikere på tværs af alle, hvortil den mest markante outperformance forekommer i kategorien *lav*, hvilket ligeledes var tilfældet ved forecast horisonten på et kvartal. Endvidere opnår modellen en forecast præcision, der er henholdsvis 58,76% og 58,20% bedre end analytikere i *lav* kategorien. Ydermere opnår modellen en stabil outperformance af analytikere på tværs af de tre andre kategorier, der er i spændet 4,74% til 31,42%, på tværs af de to forecast horisonter. Modellens outperformance af analytikere er statistisk signifikant i alle kategorier ved en forecast horisont på to og fire kvartaler, bortset fra kategorien *lav* ved en to kvartalers forecast horisont. Resultaterne vidner om, at modellens outperformance af analytikere generelt er stigende samt en højere grad af statistisk signifikans ved forlængelse af forecast horisonten, hvilket er konsistent med afhandlingens tidligere fremstillede resultater.

6.4.2 Spredning i analytikerestimater

I litteraturen er der ydermere konstateret en sammenhæng mellem analytikeres forecast præcision og spredningen i analytikerestimater. Jævnfør afsnit 2.2.2 viste resultaterne af Brown et al. (1987) en negativ relation mellem analytikeres forecast præcision og spredning i analytikerestimater. Derudover fremgik det i studiet af Ball & Ghysel (2017), at en MIDAS model outperformede analytikere ved en høj spredning. På baggrund heraf, er det fundet hensigtsmæssigt at undersøge i hvilket omfang modellens outperformance af analytikere påvirkes af graden af enighed blandt analytikere.

I forbindelse med vurdering af spredningen i analytiker estimaters betydning for modellens outperformance af analytikere, benyttes standardafvigelse som mål for spredningen i analytikerestimaterne. Endvidere er det nødvendigt at standardisere analytikeres forecast spredning, for at sikre et sammenligneligt grundlag mellem de respektive observationer. Analytikeres forecast spredning er hertil standardiseret på baggrund af aktiekursen opgjort for hver observation, med udgangspunkt i metoden fremsat af Athanassakos & Kalimipalli (2003). I tabel 6.10 illustreres modellens outperformance af analytikere opdelt i fire kategorier af analytikerspredning.

På baggrund af resultaterne fremstillet i tabel 6.10 kan det konstateres, at modellen ikke formår at outperforme analytikere i de tre kategorier *lav, mellem lav* og *mellem høj,* ved en forecast horisont på et kvartal. I disse kategorier opnår analytikere en forecast præcision, der er henholdsvis 3,45%, 6,01% og 2,72% bedre end den konstruerede machine learning model. Modellen formår dog, at outperforme analytikere markant i kategorien *høj*, hvor modellens outperformance udgør 34,79%. Endvidere kan det på baggrund af en statisk test af resultaterne konkluderes, at analytikeres outperformance af modellen er statistisk signifikant i kategorien *lav, mellem lav* og *mellem høj,* men

Tabel 6.10: Modellens outperformance af analytikere opgjort på spredning i analytiker estimater

	$h = 1$	$h = 2$	$h = 4$
	Outperformance %	Outperformance %	Outperformance %
Lav	-3,45*	5,84	28,77*
Mellem lav	-6,01**	3,65**	7,87**
Mellem høj	-2,72*	4,04**	9,84**
Høj	34,79	35,49	42,75**

Note: Tabellen angiver modellens procentuelle outperformance af analytikere i forskellige kategorier af spredning i analytikerestimater, opgjort på baggrund af MAE. Kategorierne er opdelt på baggrund af kvartilerne ved en forecast horisont på et kvartal. Ved en forecast horisont på et kvartal indikerer **, * hvorvidt analytikeres outperformance af modellen er statstisk signifikant forskellig fra 0 på et 1% og 5% niveau. Ved en forecast horisont på to og fire kvartaler indikerer **, * hvorvidt modellens outperformance er statistisk signifikant på et 1% og 5% niveau.

ikke i kategorien *høj.* Det kan derfor ikke afvises, at modellens og analytikeres forecast præcision er identisk i kategorien *høj,* ved en forecast horisont på 1 kvartal. Modsat resultaterne fremstillet ved en forecast horisont på et kvartal, er modellens outperformance mere dominerende ved en forecast horisont på henholdsvis to og fire kvartaler. Modellen formår ved disse forecast horisonter, at outperforme analytikere i alle kategorier. Endvidere er det særligt bemærkelsesværdigt, at modellens outperformance er mest markant i kategorierne *lav* og *høj*, hvilket indikerer at modellens outperformance er størst, når der er henholdsvis lav spredning og høj spredning i analytiker estimater. En lav spredning indikerer alt andet lige en større enighed i EPS estimatet blandt analytikere, hvorfor det endvidere kan virke ulogisk at modellens outperformance er så høj i kategorien *lav*. Det bør hertil bemærkes, at disse resultater kan være påvirket af testsættets komposition, herunder industrier der indgår og analytikerdækningen.

6.5 Robusthed af resultater

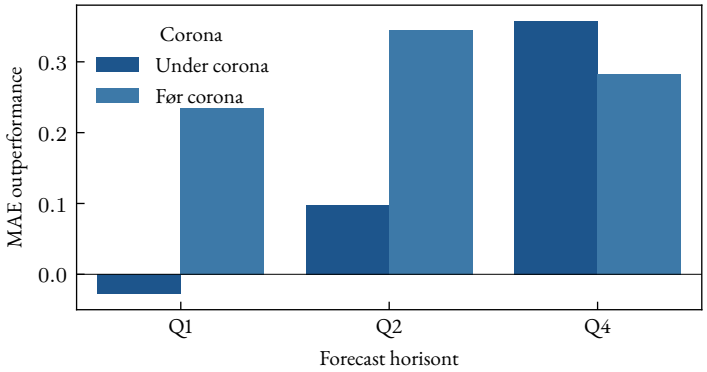
I følgende afsnit vil robustheden af resultaterne opnået i afhandlingen evalueres, med henblik på at vurdere om den konstruerede machine learning model er robust overfor ændringer testperioden, samt at analysere hvilken indflydelse længden af træningsperioden har for modellens outperformance af analytikere. Ydermere vil Corona krisens indflydelse på modellens resultater undersøges, med henblik på at evaluere modellens performance i tilfælde af en Black Swan begivenhed.

6.5.1 Corona krisen

Som belyst i afsnit 5 er resultaterne fremstillet i afhandlingen baseret på testsættet, bestående af observationer i perioden 2018 til 2020. Dette betyder endvidere, at omtrent halvdelen af testperioden var under Corona krisen. Da Ccorona krisen har vist sig at have store økonomiske implikationer for selskaber på globalt plan, og herunder også i USA, er det relevant at undersøge

hvilken betydning dette har for den konstruerede machine learning model samt analytikere. Det er hertil relevant at identificere, hvorvidt modellen har været i stand til at opfange de strukturelle ændringer i selskabers indtjening i samme omfang som analytikere, og dermed vurdere modellens anvendelighed i forbindelse med økonomiske kriser. I figur 6.3 er modellens outperformance af analytikere opgjort for henholdsvis perioden før og perioden efter corona krisen på tværs af de forskellige forecast horisonter.

Figur 6.3: Outperformance før/under corona



Note: Figuren illustrerer modellens procentuelle outperformance af analytikere før og under corona, ved forecast horisonter på et, to og fire kvartaler.

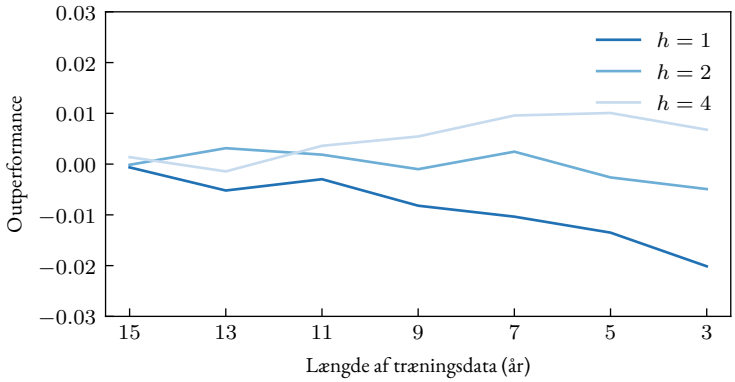
På baggrund af resultaterne fremstillet i figur 6.3 fremgår det, at Corona har haft en markant indflydelse på modellens outperformance af analytikere. I perioden før Corona formår modellen at outperforme analytikere med 23,48%, ved en forecast horisont på et kvartal. Til sammenligning, formår modellen ikke at outperforme analytikere i perioden under Corona, da analytikere her opnår en forecast præcision der er 2,72% bedre end modellen. Hverken modellens eller analytikeres outperformance viser sig at være statisk signifikant ved forcast horisonten på et kvartal. Resultaterne ved en forecast horisont på henholdsvis to og fire kvartaler vidner om en dominerede outperformance af analytikere både før og under corona. Modellen opnår hertil en outperformance af analytikere, der er statistisk signifikant på begge forecast horisonter både i perioden før og under corona. Resultaterne indikerer ydermere, at modellens outperformance er mest stabil i perioden før corona, hvor outperformance er i intervallet 23,48% til 34,5%, mens outperformance er i intervallet -2,72% til 35,75% under corona.

Endvidere kan det faktum, at modellen ikke formår at outperforme analytikere under corona ved en forecast horisont på et kvartal, samt at modellen opnår en lav outperformance ved en forecast horisont på to kvartaler sammenlignet med før corona, muligvis forklares af modellens begrænsede inputvariable i form af regnskabstal og aktiedata. Resultaterne kan derfor tænkes at indikere, at analytikeres informationsmæssige fordel er særlig fordelagtig i tilfælde af ekstreme begivenheder, samt at denne fordel mindskes ved forlængelse af forecast horisonten. Overordnet set vidner resultaterne fremstillet i figur 6.3 om modellens robusthed samt alsidighed i forbindelse med økonomiske chok, da modellen outperformer analytikere både i perioden før og under Corona med undtagelse af en forecast horisont på et kvartal under Corona, hvor modellens forecast præcision dog kun er marginalt dårligere end analytikernes.

6.5.2 Længde af træningsdata

For at undersøge modellens robustheden overfor ændringer i længden af træningsperioden simuleres syv nye versioner af den konstruerede machine learning model, med forskellige længder af træningsdata. De nye versioner af modellen vil bestå af træningsdata på henholdsvis 15, 13, 11, 9, 7, 5 og 3 års data. Fælles for alle versioner af modellen er, at testsættet, hvorudfra modellernes forecast præcision evalueres, udgør en periode 2 år, hvilket var udgangspunktet i den oprindelige model. For at belyse modellens robusthed, er den procentvise forskel sammenlignet med den oprindelige model beregnet for hver af de syv forskellige versioner af machine learning modellen. I figur 6.4 illustreres ændringen i de forskellige versioners forecast præcision sammenlignet med den oprindelige model, opgjort ved performance målet MAE.

Figur 6.4: Robusthedstjek af ændringer i længden af træningsdata



Note: Figuren illustrerer modellens outperformance af den oprindelige model, ved forskellige længder af træningsdata.

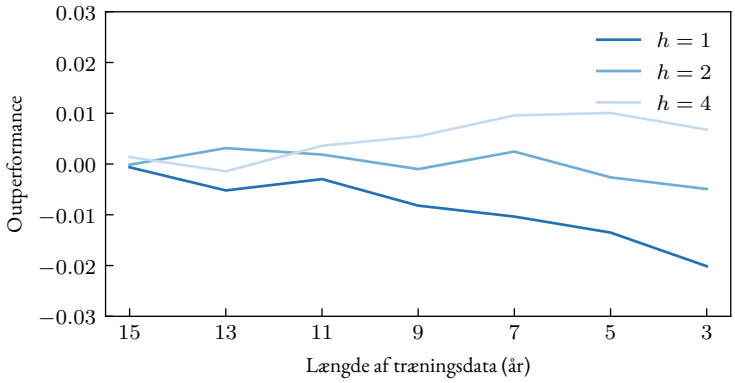
På baggrund af figur 6.4 fremgår det, at modellen er robust overfor ændringer i længden af træningsperioden på alle tre forecast horisonter, da ændringen i forecast præcision sammenlignet med den oprindelige model befinder sig i intervallet $\pm 2\%$. Det fremgår endvidere, at modellen er mest følsom overfor ændringer i længden af træningsperioden ved en forecast horisont på et kvartal, hvortil det største fald i modellens forecast præcision kan konstateres, svarende til en forringelse af forecast præcision på omtrent 2%. Modsat forbedres modellens forecast præcision ved en forkortelse af træningsperioden ved en forecast horisont på fire kvartaler, mens modellens forecast præcision ved en forecast horisont på to kvartaler er stort set uændret ved alle længder af træningsdata. Forbedringen i forecast præcision ved forkortelse af længden på træningsperioden, der gør sig gældende ved en forecast horisont på fire kvartaler kan indikere, at forecast på længere forecast horisonter baserer sig på mere fundamentale forhold, hvortil en forlængelse af træningsperioden ville medføre at modellen fokuserer for meget på kortsigtede forhold.

Overordnet set kan det konkluderes at den konstruerede machine learning model er robust overfor ændringer i længden af træningsperioden, hvilket indikerer at modellen er god til at tilpasse sig input data. Samtidig indikerer resultaterne, at modellen har stor anvendelighed selv i tilfælde, hvor mængden af træningsdata er begrænset. Dette er særligt relevant i forbindelse med en praktisk implementering af modellen, da det understreger modellens evne til, at trænes på en anden træningsperiode end perioden benyttet i afhandlingen.

6.5.3 Rolling træning test

Afsnit 6.5.1 og 6.5.2 kunne konstatere, at modellen er relativ robust overfor en black swan begivenhed, samt at længden af træningsperioden har en marginal indflydelse på modellens forecast præcision. Endvidere er det ligeledes relevant at undersøge modellens robusthed i forbindelse med ændring af testperioden. For at undersøge dette udarbejdes et rolling trænings- og testsæt, hvor trænings- og testperioden løbende forskydes med to år. Endvidere er seks nye trænings- og testsæt genereret, og resultaterne heraf fremgår i figur 6.5, hvor modellens outperformance af analytikere er opgjort for alle tre forecast horisonter.

Figur 6.5: Robusthedstjek af rolling træning/test



Note: Figuren illustrerer antallet af imputede værdier fordelt på de forskellige inputvariable.

På baggrund af figur 6.5 kan det konstateres at outperformance falder markant under finanskrisen, hvor analytikernes outperformance på de tre forecast horisonter er henholdsvis 15%, 12% og 1%. Dette er i kontrast til robustheden under Corona, hvor det blev identificeret at modellen outperformance analytikere under Corona på ved en forecast horisont på to og fire kvartaler. Årsager til modellens markant dårligere performance sammenlignet med analytikere under finanskrisen kan skyldes de mere fundamentale karakteristika ved finanskrisen, der i højere grad gør det muligt for analytiker at benytte deres informationsfordel.

Yderligere kan det på baggrund af figuren konstateres, at de forskellige testperioders resultater er konsistent med tidligere resultater fremstillet i afhandlingen, da modellens outperformance af analytikere er skiftende ved en forecast horisont på et kvartal, hvilket understøtter den statistisk insignifikans identificeret i afsnit 6.1. Modsat indikerer figuren, at modellen outperformer analytikere under alle testperioder, med undtagelse af finanskrisen, ved en forecast horisont på henholdsvis to og fire kvartaler, hvilket er konsistent med den statistiske signifikans identificeret i afsnit 6.1. På baggrund af rolling træning og test kan det dermed konkluderes, at den udviklede machine learning model er robust overfor ændringer i testperioden, og dertil kan anvendes på andre perioder end den undersøgte testperiode i afhandlingen.

7 Diskussion

I dette kapitel foretages en diskussion af metode og resultater fremstillet i afhandlingen. I kapitlet vil begrænsninger samt metodiske valg, der kan tænkes at have indflydelse på afhandlingens resultater diskuteres. Afslutningsvis vil afhandlingens resultater relateres til tidligere litteratur med henblik på at klarlægge afhandlingens bidrag til forskningsfeltet.

Key takeaways: En af de væsentligste begrænsninger i afhandlingen er, at machine learnings anvendeligheden vurderes på baggrund af en enkeltstående machine learning model fremfor flere. Afhandlingens resultater adskiller sig fra tidligere studier ved at påvise, at en machine learning model er i stand til at outperforme analytikere ved en kort forecast horisont. Afhandlingens resultater er konsistent med tidligere studier, der ligeledes konstaterede en høj anvendelighed af automatiserede forecast modeller ved forecast af mindre selskaber, samt i tilfælde med en lav analytikerdækning og en høj spredning i analytiker estimater.

7.1 Begræsninger

I afhandlingen beror vurderingen af machine learnings anvendelighed i forbindelse med forecast af EPS udelukkende på én enkelt machine learning algoritmes resultater. Endvidere er modelvalget i afhandlingen begrænset til Catboost algoritmen, da denne algoritme viste sig at opnå den bedste forecast præcision på valideringssættet, med udgangspunkt i en basemodel på en et kvartals forecast horisont. Da valget af algoritme er baseret på basemodellens performance, er det muligt, at andre algoritmer, kunne have opnået en bedre performance end Catboost modellen efter foretagelse af feature engineering og feature selektion samt hyperparameter. Jævnfør afsnit 5.3 fremgik det blandt at LightGBM algoritmen opnåede en marginalt lavere forecast præcision end Catboost algoritmen, hvorfor det ikke kan afvises,at denne potentielt kunne outperforme den valgte Catboost algoritme.I litteraturen benyttes flere forskellige machine learning algoritmer til at evaluere machine learnings performance sammenligne med analytikere, herunder i studiet af Hunt et al. (2019), hvoriblandt algoritmerne random forrest, SVM og gradient boosting[13] anvendes. Særligt i forbindelse med konstruering af industrispecifikke modeller kunne inddragelsen af forskellige algoritmer være relevant, da dette kunne have ændret konklusionen om, hvorvidt en generel eller en industri specifik model var mest optimal i den givne industri.

I afhandlingen sammenlignes analytikeres og den konstruerede machine learning models forecast præcision på tre forskellige forecast horisonter, der baserer sig på kvartalsmæssige opgørelser. Endvidere kunne det have givet et mere dybdegående indblik i forecast præcisionen på tværs af forecast horisonter, såfremt forecast præcision var opgjort på månedlig basis. Blandt andet Bradshaw et al. (2012) analyserer analytikeres forecast præcision opgjort på månedlig basis sammenholdt med

[13] Forfatteren nævner ikke hvilke gradient boosting algoritmer der benyttes eksplicit.

et random walk forecast. Grundet begrænset adgang til IBES databasen, der gennem CBS licens kun var muligt at tilgå via Eikon terminal, var det ikke muligt at tilgå analytikeres konsensus estimater opgjort på månedlig basis. Da vi i afhandlingen ønsker at undersøge, hvorvidt det er muligt at outperforme analytikere, er det essentielt at der forelægger analytiker estimater. Af denne årsag var det kun var muligt at analysere forecast præcision aggregeret på kvartalsniveau, grundet CBS begrænsede API adgang til Eikon.

7.2 Refleksion

Ved fortolkning af de fremstillede resultater i afhandlingen er det vigtigt at bemærke, at analytikeres forecast præcision er baseret på analytikeres konsensusestimat, og dermed udgør en gennemsnitsbetragtning af de respektive analytikeres forecast. Resultaterne fremstillet i afhandlingen angiver dermed, hvorvidt den konstruerede machine learning model var i stand til at opnå en højere forecast præcision end analytikere i gennemsnit. Dette betyder endvidere at det potentielt er muligt, at individuelle analytikere kan have opnået en bedre forecast præcision end modellen til trods for, at det i afhandlingen konkluderes, at modellen outperformer analytikere. Denne valgte måde at afspejle analytikeres forecast præcision, er konsistent med litteraturens måde af afspejle analytikeres forecast præcision og stemmer samtidig overens med, hvordan praktikere benytter analytiker estimater. Endvidere er det i litteraturen konstateret, at visse analytikere konsekvent opnår en højere forecast præcision i deres indtjeningsestimater sammenlignet med deres peers. Både Kerl & Ohlert (2014) og Stickel (1992) konkluderede at højt rangerede analytikere konsekvent outperformede analytikere, der ikke var højt rangeret, hvilket dermed indikerer at visse analytikere er systematisk bedre til at forecaste. På baggrund af de høj rangerede analytikeres systematiske evne til at outperforme peers, er det muligt at en sammenligning af den konstruede machine learning models og højt rangerede analytikeres forecast præcision kunne ændre afhandlingens konklusioner. Det ville endvidere være muligt at foretage denne sammenligning på baggrund af Starmine Smart Estimates estimatet fra Refinitiv (Refinitiv, 2021).

7.3 Relater til tidligerelitteratur

Analysens overordnede resultater viste, at det var muligt at konstruere en machine learning model med regnskabs- og aktiedata som primært inputvariable, der var i stand til at outperforme analytikere, ved en forecast horisont på et, to og fire kvartaler. Disse resultater adskiller sig fra resultaterne fremstillet af eksempelvis Bradshaw et al. (2012), Collings & Hopwood (1980) og Fried & Givoly (1982) da disse studier konkluderede, at det ikke var muligt for automatiserede forecast modeller at outperforme analytikere, ved en forecast horisont en kort forecast horisont.

Resultaterne er dog konsistente med nyere studier, der gør brug af machine learning, herunder Van Binsbergen et al. (2020) og Elend et al. (2020), hvis modeller formåede at outperforme analytikere, ved en forecast horisont på et kvartal. Afhandlingens resultater ved forecast horisonter på to og fire kvartaler er konsistente med resultaterne i litteraturen, der ligeledes identificerede en forringelse i modeller samt analytikeres forecast præcision ved forlængelse af forecast horisonten (e.g. Brown, Richardson, et al., 1987; Kross et al., 1990; Lys & Soo, 1995; van Binsbergen et al., 2020). Ydermere er resultaterne på to og fire kvartalers forecast horisont konsistente med resultaterne fremstillet af

Fried & Givoly (1982) og Bradshaw et al. (2012), der ligeledes indikerede, at analytikere havde svært ved outperforme random walk modeller på længere forecast horisonter.

En dekomponering af modellens og analytikeres forecast præcision på tværs af forskellige industrier viste endvidere at modellen formåede at modellen formåede at outperforme analytikere i størstedelen af de 11 industrier, på tværs af de tre forecast horisonter. Modsat Ball & Ghysels (2017), der ved en forecast horisont på et kvartal, formåede at opnå en statistisk signifikant outperformance af analytikere i fem industrier, var modellen i afhandlingen kun statisk signifikant i én industri. Endvidere er variationen i henholdsvis modellens og analytikeres forecast præcisionen på tværs af industrier konstateret i afhandlingen konsistent med Ball & Ghysels (2017), der ligeledes kunne konstatere væsentlige forskelle på tværs af industrier. Resultaterne i afhandlingen vidnede om, at det ikke var muligt at outperforme analytikere i Finance industrien, hvilket stemmer overens med resultaterne fremstillet i litteratur, hvor Elend et al (2020) deep learning model opnåede en dårligere performance på finansielle virksomheder sammenlignet med andre industrier. Disse resultater er endvidere konsistente med Ou and Penman (1989), der valgte at ekskludere finansvirksomheder i datagrundlaget, grundet den anderledes klassificering af regnskabsposter og betydningen af regnskabsnøgletal. Denne forskel kan tænkes at forklare, hvorfor det ikke var muligt for den konstruerede machine learning model at outperfome analytikere i *Finance* industrien i afhandlingen.

På baggrund af en dekomponering af resultaterne i virksomhedstørrelse kunne det i afhandlingen konstateres at modellen konsekvent outperformede analytikere på tværs af forecast horisonterne, ved forecast af små selskabers EPS. Ydermere indikerede resultaterne at der forekom en positiv sammenhæng mellem virksomhedsstørrelsen og analytikeres outperformance af modellen. Resultaterne i afhandlingen er endvidere konsistent med litteraturen, hvor af Ball & Ghysels (2017), ligeledes konstaterede en større outperformance af analytikere ved forecast af mindre selskaber. Ydermere fremgik det i studiet af Brown et al. (1987). Afhandlingens resultater er ligeledes konsistente med resultaterne af Brown et al. (1987), der konstaterede at analytikere outperformede tidseriemodeller ved forecast af store virksomheder, samt at en kombination af tidseriemodeller og analytikeres var særlig relevant ved forecast af mindre virksomheders EPS. Resultaterne dekomponeret på analytikerdækning viste, at modellen opnåede en markant outperformance af analytikere ved en lav analytikerdækning. Disse resultater er konsistente med litteraturen, hvor Lang & Lundholm (1996) og Alford & Berger (1999) konstaterede en positiv korrelation mellem antallet af analytikere og forecast præcisionen. Lang & Lundholm (1996) og Alford & Berger (1999) argumenterede hertil for, at årsagen til denne sammenhæng var, at hver analytiker bidrager med ny information, der forbedrer konsensusestimatet.

I afhandlingen kunne det på baggrund af spredningen i analytikerestimater konstateres, at modellens opnåede en markant bedre forecastpræcision end analytikere i tilfælde, hvor der forekom en høj spredning i analytikerestimatet. Disse resultater er endvidere konsistente med Brown et al. (1987) resultater, der ligeledes viste at brugen af tidsseriemodeller var særligt brugbare ved heterogene forventninger hos analytikerne. I det nyere studie af Ball & Ghysels (2017) viste resultaterne ligeledes at deres konstruerede MIDAS model performede bedre end analytikerne ved en høj spredning i analytikerestimatet. Modsat resultaterne fra Brown et al., (1987) og Ball & Ghysels (2017) fremgik det i afhandlingens resultater, at den konstruerede machine learning model ligeledes performede godt ved en lav analytikerspredning. Hvorvidt forskelle i resultaterne skyldes forskelle i det underlæggende datagrundlag mellem afhandlingen og disse studier er ikke muligt at verificere.

8 Konklusion og forlag til yderligere research

I det følgende kapitel vil der konkluderes på resultaterne fremstillet i afhandlingen, for at besvare den fremstillede problemformulering. Efterfølgende fremsættes forslag til yderligere forskning, der potentielt kunne forbedre resultaterne fremstillet i afhandlingen samt en diskussion af relevante områder, der er uden for scopet af afhandlingen.

8.1 Konklusion

Afhandlingen undersøger i hvilket omfang en machine learning model baseret på regnskabstal og aktiedata kan anvendes til at estimere selskabers fremtidige indtjening per aktie. På baggrund af et datagrundlag bestående af 128.668 observationer fra 3.754 selskaber i alt, var det muligt at implementere machine learning algoritmen Catboost samt at sammenholde resultaterne heraf med aktieanalytikeres forecast, for at evaluere modellens anvendelighed.

Afhandlingens resultater skaber overordnet set evidens for, at estimering af selskabers fremtidige indtjening kan automatiseres ved brug af machine learning. Resultaterne fremstillet i afhandlingen vidnede om, at forecast foretaget af en machine learning model opnåede en højere forecast præcision end analytikere på tværs af forecast horisonter på et, to og fire kvartaler. Det blev yderligere konkluderet, at modellens outperformance af analytikere steg ved forlængelse af forecast horisonten, samt at denne outperformance var statistisk signikant ved forecast horisonter på to og fire kvartaler. En dekomponering af machine learning modellens resultater på baggrund af 11 industrier vidnede om, at det var mest hensigtsmæssigt at implementere en generel model fremfor industrispecifikke modeller i størstedelen af industrierne. Resultaterne heraf indikerede, at relevansen af industrispecifikke modeller i højere grad ville komme til udtryk ved inkorporering af industrispecifikke inputvariable. På baggrund af en sammenligning af machine learning modellens og analytikeres resultater kan det konkluderes, at modellen outperformede analytikere på tværs af industrier, herunder særligt i industrierne *Healthcare*, *Energy* og *Chemicals & allied products*. Ydermere kan det konkluderes, at machine learning modellen er mindre anvendelig i industrierne *Finance* og *Telecom*, hvor analytikere konsekvent outperformer modellen. På baggrund af afhandlingens resultater kan det konkluderes, at machine learning modellens anvendelighed i høj grad påvirkes af selskabers størrelse og analytikerdækning. En dekomponering af machine learning modellens forecast præcision opgjort på baggrund af virksomhedsstørrelse viste, at modellen var særligt anvendelig i forbindelse med forecast af mindre selskabers EPS, hvor modellen konsekvent outperformede analytikere. Resultaterne opgjort på baggrund af analytikerdækning samt spredningen i analytikerestimater viste, at modellen performede særdeles godt i tilfælde med en lav analytikerdækning samt høj spredning i estimaterne blandt analytikere.

Et robusthedstjek af machine learning modellens resultater viste, at modellens performance er robust over for ændringer i længden af træningsperioden samt valget af testperiode. Resultaterne understreger endvidere machine learning modellens alsidighed og anvendelighed i forbindelse med en begrænset træningsperiode samt under både økonomiske opsving og kriser. Afhandlingens resultater understreger potentialet for anvendelsen af machine learning i forbindelse med estimering af selskabers fremtidige indtjening. Afhandlingen udfordrer dermed den hidtidige opfattelse af, at analytikeres estimater udgør den bedste proxy for selskabers fremtidige indtjening, og skaber dermed et fundament for yderligere research.

8.2 Forslag til yderligere research

Afhandlingens resultater vidnede om machine learnings lovende resultater i forbindelse med automatisering af selskabers indtjeningsestimater. Grundet ønsket om at konstruere en simpel machine learning model, var det primære datagrundlag i afhandlingen begrænset til regnskabstal og aktiedata, hvortil det kunne være interessant for fremtidig research at inkludere yderligere data. Der findes adskillelige inputvariable, der potentielt kunne forbedre machine learning modellens forecast præcision, heriblandt industrispecifik- og makroøkonomiskinformation, der i litteraturen er identificeret som den primære årsag til analytikeres overlegne performance (Fried & Givoly, 1982; O'brien, 1988). Inddragelsen af disse informationer kunne potentielt minimere analytikeres informationsfordel, der særligt kom til udtryk under økonomiske kriser samt inden for specifikke industrier. Ydermere ville anvendelsen af Natural Language Processing (NLP) kunne benyttes til at opfange positive og negative signaler i medierne, der kunne være indikativ for selskabers fremtidige indtjening. Babii et al. (2020) viste lovende resultater for brugen af NLP i forbindelse med fastlæggelse af price-earnings multipel.

Set i lyset af MIFID II reguleringen på det europæiske marked, der de seneste år har medført et stort fald i analytikerdækningen i Europa, vil det være relevant for yderligere research at efterprøve afhandlingens resultater på det europæiske marked (Lang et al., 2019). Da det ikke kan afvises, at der forekommer fundamentale forskelle i regnskabstal og aktiedatas evne til at forklare selskabers fremtidige indtjening på henholdsvis det amerikanske og europæiske marked, er det dermed ikke muligt at generalisere afhandlingens konklusioner til det europæiske marked. Ydermere kan forskellen i regnskabspraksis samt kvaliteten af analytikeres estimater have indflydelse på machine learnings anvendelse i Europa, hvorfor en klarlæggelse af anvendeligheden af machine learning i forbindelse med estimering af europæiske selskabers EPS kræver yderligere research.

A Inputvariable

A.1 Data variable før cleaning

Tabel A.1: Datavariable før cleaning

Compustat	IBES Actuals	IBES summary	CRSP	Fama Fren
GVKEY	Instrument	Instrument	date	Date
datadate	Earnings Per Share Reported - Actual	Calc Date	PERMNO	NoDur
fyearq	Calc Date	Period End Date	TICKER	Durbl
fqtr	Period End Date	Earnings Per Share Reported - Mean	CUSIP	Manuf
fyr		Earnings Per Share Reported - Standard Deviation	PRC	Enrgy
indfmt		Earnings Per Share Reported - High	VOL	Chems
consol		Earnings Per Share Reported - Low	RET	BusEq
popsrc			SHROUT	Telcm
datafmt			CFACPR	Utils
tic			CFACSHR	Shops
cusip			ALTPRC	Hlth
conm			ALTPRCDT	Money
curcdq			adj_close	Other
curncdq			count	
datacqtr			count_obs	
datafqtr			adj_return	
rp				
rdq				
apq				
atq				
ceqq				
cogsq				
dpq				
gdwlq				
niq				
nopiq				
oiadpq				
piq				
ppentq				
rcpq				
rectq				
revtq				
txtq				
xoprq				
xrdq				
xsgaq				
costat				
dvpspq				
dvpsxq				
mkvaltq				

loc
sic
perm

A.2 Baseline inputvariable

Tabel A.2: Inputvariable i baseline model

Input Variable
Accounts_payable(Apq)
COGS(Cogsq)
Common_equity(Ceqq)
DPS_ex_date(dvpsxq)
Deprect_and_ammor(Dpq)
EPS_lagged
GVKEY
Income taxes_total(Txtq)
Non_operating_income(Nopiq)
Operating_income_after_deprec(Oiadpq)
Opex_total(Xoprq)
PP&E_total_net(Ppentq)
Pretax_income(Piq)
Receviables(Rectq)
Revenue_total(Revtq)
adj_return
industry_fama
market_cap
niq
merafkast
ALTPRC
month
year
Div_yield

Note: Baseline modellen anvender default modelparametre fra Catboost.

A.3 Endelige inputvariable

Tabel A.3: Inputvariable i den endelige model

Input variable	Input variable
EPS_actual	Operating_income_after_deprec(Oiadpq)_1
Accounts_payable(Apq)	Operating_income_after_deprec(Oiadpq)_2
COGS(Cogsq)	Operating_income_after_deprec(Oiadpq)_3
Common_equity(Ceqq)	Operating_income_after_deprec(Oiadpq)_4
DPS_ex_date(dvpsxq)	Operating_income_after_deprec(Oiadpq)_5
Deprect_and_ammor(Dpq)	Opex_total(Xoprq)_1
EPS_lagged	Opex_total(Xoprq)_2
GVKEY	Opex_total(Xoprq)_3
Income taxes_total(Txtq)	Opex_total(Xoprq)_4
Non_operating_income(Nopiq)	Opex_total(Xoprq)_5
Operating_income_after_deprec(Oiadpq)	PP&E_total_net(Ppentq)_1
Opex_total(Xoprq)	PP&E_total_net(Ppentq)_2
PP&E_total_net(Ppentq)	PP&E_total_net(Ppentq)_3
Pretax_income(Piq)	PP&E_total_net(Ppentq)_4
Receviables(Rectq)	PP&E_total_net(Ppentq)_5
Revenue_total(Revtq)	Pretax_income(Piq)_1
adj_return	Pretax_income(Piq)_2
industry_fama	Pretax_income(Piq)_3
market_cap	Pretax_income(Piq)_4
niq	Pretax_income(Piq)_5
merafkast	Receviables(Rectq)_1
ALTPRC	Receviables(Rectq)_2
month	Receviables(Rectq)_3
year	Receviables(Rectq)_4
Div_yield	Receviables(Rectq)_5
Accounts_payable(Apq)_1	Revenue_total(Revtq)_1
Accounts_payable(Apq)_2	Revenue_total(Revtq)_2
Accounts_payable(Apq)_3	Revenue_total(Revtq)_3
Accounts_payable(Apq)_4	Revenue_total(Revtq)_4
Accounts_payable(Apq)_5	Revenue_total(Revtq)_5
COGS(Cogsq)_1	adj_return_1
COGS(Cogsq)_2	adj_return_2
COGS(Cogsq)_3	adj_return_3
COGS(Cogsq)_4	adj_return_4
COGS(Cogsq)_5	adj_return_5
Common_equity(Ceqq)_1	market_cap_1
Common_equity(Ceqq)_2	market_cap_2
Common_equity(Ceqq)_3	market_cap_3
Common_equity(Ceqq)_4	market_cap_4
Common_equity(Ceqq)_5	market_cap_5
DPS_ex_date(dvpsxq)_1	niq_1

Tabel A.3: Inputvariable i den endelige model

Input variable	Input variable
DPS_ex_date(dvpsxq)_2	niq_2
DPS_ex_date(dvpsxq)_3	niq_3
DPS_ex_date(dvpsxq)_4	niq_4
DPS_ex_date(dvpsxq)_5	niq_5
Deprect_and_ammor(Dpq)_1	merafkast_1
Deprect_and_ammor(Dpq)_2	merafkast_2
Deprect_and_ammor(Dpq)_3	merafkast_3
Deprect_and_ammor(Dpq)_4	merafkast_4
Deprect_and_ammor(Dpq)_5	merafkast_5
EPS_lagged_1	ALTPRC_1
EPS_lagged_2	ALTPRC_2
EPS_lagged_3	ALTPRC_3
EPS_lagged_4	ALTPRC_4
EPS_lagged_5	ALTPRC_5
Income taxes_total(Txtq)_1	Div_yield_1
Income taxes_total(Txtq)_2	Div_yield_2
Income taxes_total(Txtq)_3	Div_yield_3
Income taxes_total(Txtq)_4	Div_yield_4
Income taxes_total(Txtq)_5	Div_yield_5
Non_operating_income(Nopiq)_1	Non_operating_income(Nopiq)_4
Non_operating_income(Nopiq)_2	Non_operating_income(Nopiq)_5
Non_operating_income(Nopiq)_3	

B SIC KODER

Tabel B.1: Beskrivelse af industrier

Industri	SIC	Beskrivelse
Consumer Nondurables	0100-0999,2000-2399,2700-2749, 2770-2799,3100-3199,3940-3989	Udvalgsvarer
Consumer Durables	2500-2519,2590-2599,3630-3659,3710-3711, 3714-3714,3716-3716,3750-3751,3792-3792, 3900-3939,3990-3999	Forbrugsvarer
Manufacturing	2520-2589,2600-2699,2750-2769,3000-3099, 3200-3569,3580-3629,3700-3709,3712-3713, 3715-3715,3717-3749,3752-3791,3793-3799, 3830-3839,3860-3899	Produktion
Energy	1200-1399,2900-2999	Energi (olie, gas mfl.)
Chemicals & A.P	2800-2829,2840-2899	Kemikaler mm.
Business equipment	3570-3579,3660-3692,3694-3699,3810-3829, 7370-7379	Elektronik, computere og software
Telecom	4800-4899	Teleselskaber
Wholesale/Retail	5000-5999,7200-7299,7600-7699	Engros- og detailsalg
Healthcare	2830-2839,3693-3693,3840-3859,8000-8099	Medicinal industrien
Finance	6000-6999	Finansielle virksomheder
Other	Andre	Andre

C Analyse

C.1 Marketcap uden Healthcare

Tabel C.1: Forecast performance market cap renset for *Healthcare*

	h					
	h=1		h=2		h=4	
	Analytiker	Model	Analytiker	Model	Analytiker	Model
Small cap	1,15	1,09	1,15	1,07	1,13	1,08
Mid cap	1,24	1,09	1,56	1,11	1,46	1,19
Large cap	1,31	1,14	1,41	1,25	1,77	1,42

Note: Tabellen angiver modellen og analytikeres forecast performance i forskellige market cap kategorier, opgjort ved performance målet RMSE. Resultaterne er renset for observationer fra industrien *Healthcare* ved en forecast horisont på et, to og fire kvartaler.

D Python

Github Repo: Python kan ses direkte på Github Tryk her
Backup link: https://github.com/fc0712/Speciale-Python

Jupyter notebook viewer: Der kan med fordel benyttes Jupyter Notebook viewer til at se Python notebooks: Tryk her
Backup link: https://nbviewer.jupyter.org/github/fc0712/Speciale-Python/tree/main/

Bibliografi

Abarbanell, J. & Lehavy, R. (2003). Biased forecasts or biased earnings? The role of reported earnings in explaining apparent bias and over/underreaction in analysts' earnings forecasts. *Journal of Accounting and Economics*, *36*(1-3), 105–146. DOI: 10.1016/j.jacceco.2003.11.001

Abhigyan. (2020). *Feature Selection For Dimensionality Reduction(Embedded Method)*. Hentet 22. marts 2021, fra https://medium.com/analytics-vidhya/feature-selection-for-dimensionality-reduction-embedded-method-e05c74014aa

Akiba, T., Sano, S., Yanase, T., Ohta, T. & Koyama, M. (2019). Optuna: A Next-Generation Hyperparameter Optimization Framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623–2631. DOI: 10.1145/3292500.3330701

Alghofaili, Y., Albattah, A. & Rassam, M. A. (2020). A Financial Fraud Detection Model Based on LSTM Deep Learning Technique. *Journal of Applied Security Research*, *15*(4), 498–516. DOI: 10.1080/19361610.2020.1815491

Amel-Zadeh, A., Calliess, J.-P., Kaiser, D. & Roberts, S. (2020). Machine Learning-Based Financial Statement Analysis. *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3520684

Andreou, E., Ghysels, E. & Kourtellos, A. (2013). Should Macroeconomic Forecasters Use Daily Financial Data and How? *Journal of Business & Economic Statistics*, *31*(2), 240–251. DOI: 10.1080/07350015.2013.767199
_eprint: https://doi.org/10.1080/07350015.2013.767199

Apple Inc. (2019, 30. juli). *Apple Reports Third Quarter Results*. Hentet 9. marts 2021, fra https://www.apple.com/newsroom/2019/07/apple-reports-third-quarter-results/

Arand, D., Kerl, A. & Walter, A. (2015). When Do Sell-side Analyst Reports Really Matter? Shareholder Protection, Institutional Investors and the Informativeness of Equity Research: When Do Sell-side Analyst Reports Really Matter? *European Financial Management*, *21*(3), 524–555. DOI: 10.1111/j.1468-036X.2013.12028.x

Arand, D. & Kerl, A. G. (2015). Sell-Side Analyst Research and Reported Conflicts of Interest. *European Financial Management*, *21*(1), 20–51. DOI: 10.1111/j.1468-036X.2012.00661.x
_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-036X.2012.00661.x

Arcadinho, S. & Mateus, P. (2019, 22. marts). *Time Series Imputation*. arXiv: 1903.09732 [cs, stat]. Hentet 8. marts 2021, fra http://arxiv.org/abs/1903.09732

Ashton, D. & Wang, P. (2013). Terminal valuations, growth rates and the implied cost of capital. *Review of Accounting Studies*, *18*(1), 261–290. DOI: 10.1007/s11142-012-9208-5

Athanassakos, G. & Kalimipalli, M. (2003). Analyst Forecast Dispersion and Future Stock Return Volatility. *Quarterly Journal of Business and Economics*, *42*(1/2), 57–78.

Athavale, M., Myring, M. & Gröber, R. (2013). Analyst Earnings Forecasts in BRIC Countries. *THE JOURNAL OF APPLIED BUSINESS AND ECONOMICS*, *14*, 71–82.

Azevedo, V., Bielstein, P. & Gerhart, M. (2021). Earnings forecasts: The case for combining analysts' estimates with a cross-sectional model. *Review of Quantitative Finance and Accounting*, *56*(2), 545–579. DOI: 10.1007/s11156-020-00902-z

Aziz, S. & Dowling, M. (2019). Machine Learning and AI for Risk Management. I T. Lynn, J. G. Mooney, P. Rosati & M. Cummins (Red.), *Disrupting Finance: FinTech and Strategy in the 21st Century* (s. 33–50). Springer International Publishing. DOI: 10.1007/978-3-030-02330-0_3

Babii, A., Ball, R. T., Ghysels, E. & Striaukas, J. (2020). Machine Learning Panel Data Regressions with an Application to Nowcasting Price Earnings Ratios. *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3670847

Bae, K.-H., Stulz, R. M. & Tan, H. (2008). Do local analysts know more? A cross-country study of the performance of local analysts and foreign analysts. *Journal of Financial Economics*, *88*(3), 581–606. DOI: 10.1016/j.jfineco.2007.02.004

Ball, R. & Brown, P. (1968). An Empirical Evaluation of Accounting Income Numbers. *Journal of Accounting Research*, *6*(2), 159–178. DOI: 10.2307/2490232

Ball, R. & Watts, R. (1972). Some Time Series Properties of Accounting Income. *Journal of Finance*, *27*(3), 663–81. Hentet 7. april 2021, fra https://econpapers.repec.org/article/blajfinan/v_3a27_3ay_3a1972_3ai_3a3_3ap_3a663-81.htm

Ball, R. T. & Ghysels, E. (2017, 22. maj). *Automated Earnings Forecasts: Beat Analysts or Combine and Conquer?* (SSRN Scholarly Paper Nr. ID 2972185). Social Science Research Network. Rochester, NY. Hentet 10. februar 2021, fra https://papers.ssrn.com/abstract=2972185

Bandyopadhyay, S. P., Brown, L. D. & Richardson, G. D. (1995). Analysts' use of earnings forecasts in predicting stock returns: Forecast horizon effects. *International Journal of Forecasting*, *11*(3), 429–445. DOI: 10.1016/0169-2070(95)00593-0

Baranes, A. & Palas, R. (2019). EARNING MOVEMENT PREDICTION USING MACHINE LEARNING-SUPPORT VECTOR MACHINES (SVM). *Journal of Management Information & Decision Sciences*, *22*(2), 36–53. http://esc-web.lib.cbs.dk/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=141174726&site=ehost-live&scope=site

Barber, B. M., Lehavy, R. & Trueman, B. (2007a). Comparing the stock recommendation performance of investment banks and independent research firms. *Journal of Financial Economics*, *85*(2), 490–517. DOI: 10.1016/j.jfineco.2005.09.004

Barber, B. M., Lehavy, R. & Trueman, B. (2007b). Comparing the stock recommendation performance of investment banks and independent research firms. *Journal of Financial Economics*, *85*(2), 490–517. DOI: 10.1016/j.jfineco.2005.09.004

Barniv, R., Myring, M. J. & Thomas, W. B. (2005). The Association between the Legal and Financial Reporting Environments and Forecast Performance of Individual Analysts*. *Contemporary Accounting Research*, *22*(4), 727–758. DOI: 10.1506/75CE-HAT9-RH17-LAD7

Barsky, R. B. & Miron, J. A. (1989). The Seasonal Cycle and the Business Cycle. *Journal of Political Economy*, *97*(3), 503–534. DOI: 10.1086/261614

Barth, M. E. & Kallapur, S. (1996). The Effects of Cross-Sectional Scale Differences on Regression Results in Empirical Accounting Research*. *Contemporary Accounting Research*, *13*(2), 527–567. DOI: 10.1111/j.1911-3846.1996.tb00514.x
_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1911-3846.1996.tb00514.x

Basu, S. & Markov, S. (2004). Loss function assumptions in rational expectations tests on financial analysts' earnings forecasts. *Journal of Accounting and Economics*, *38*, 171–203. DOI: 10.1016/j.jacceco.2004.09.001

Bathke, A. W. & Lorek, K. S. (1984). The Relationship between Time-Series Models and the Security Market's Expectation of Quarterly Earnings. *The Accounting Review*, *59*(2), 163–176.

Bauer, J., Angelini, O. & Denev, A. (2017). Imputation of Multivariate Time Series Data - Performance Benchmarks for Multiple Imputation and Spectral Techniques. *SSRN Electronic Journal*. DOI: 10.2139/ssrn.2996611

Bayen, A. M. & Siauw, T. (2015). Interpolation. *An Introduction to MATLAB® Programming and Numerical Methods for Engineers* (s. 211–223). Elsevier. DOI: 10.1016/B978-0-12-420228-3.00014-2

Beaulieu, J. J., MacKie-Mason, J. K. & Miron, J. A. (1992). Why Do Countries and Industries with Large Seasonal Cycles Also Have Large Business Cycles? *The Quarterly Journal of Economics*, *107*(2), 621–656. DOI: 10.2307/2118484

Beaver, W., Lambert, R. & Morse, D. (1980). The information content of security prices. *Journal of Accounting and Economics*, *2*(1), 3–28. DOI: 10.1016/0165-4101(80)90013-0

Beaver, W. H., Lambert, R. A. & Ryan, S. G. (1987). The information content of security prices. *Journal of Accounting and Economics*, *9*(2), 139–157. DOI: 10.1016/0165-4101(87)90003-6

Beckers, S., Steliaros, M. & Thomson, A. (2004). Bias in European Analysts' Earnings Forecasts. *Financial Analysts Journal*, *60*(2), 74–85.

Behn, B. K., Choi, J.-H. & Kang, T. (2007). Audit Quality and Properties of Analyst Earnings Forecasts. *SSRN Electronic Journal*. DOI: 10.2139/ssrn.1013214

Bernhardt, D., Campello, M. & Kutsoati, E. (2006). Who herds? *Journal of Financial Economics*, *80*(3), 657–675. DOI: 10.1016/j.jfineco.2005.07.006

Bhojraj, S., Charles M. C. Lee & Oler, D. K. (2003). What's My Line? A Comparison of Industry Classification Schemes for Capital Market Research. *Journal of Accounting Research*, *41*(5), 745–774. Hentet 2. maj 2021, fra http://www.jstor.org.esc-web.lib.cbs.dk/stable/3542375

Bikhchandani, S., Hirshleifer, D. & Welch, I. (1992). A Theory of Fads, Fashion, Custom, and Cultural Change in Informational Cascades. *Journal of Political Economy*, *100*(5), 992–1026. Hentet 26. januar 2021, fra https://econpapers.repec.org/article/ucpjpolec/v_3a100_3ay_3a1992_3ai_3a5_3ap_3a992-1026.htm

Boehmke, B. & Greenwell, B. (2020). *Hands-On Machine Learning with R*. Hentet 20. april 2021, fra https://bradleyboehmke.github.io/HOML/

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., Ljung, G. M. & Ljung, G. M. (2015). *Time Series Analysis : Forecasting and Control*. John Wiley & Sons, Incorporated. http://ebookcentral.proquest.com/lib/kbhnhh-ebooks/detail.action?docID=2064681

Bradshaw, M. T., Drake, M. S., Myers, J. N. & Myers, L. A. (2012). A re-examination of analysts' superiority over time-series forecasts of annual earnings. *Review of Accounting Studies*, *17*(4), 944–968. DOI: 10.1007/s11142-012-9185-8

Bradshaw, M. T. & Sloan, R. G. (2002). GAAP versus The Street: An Empirical Assessment of Two Alternative Definitions of Earnings. *Journal of Accounting Research*, *40*(1), 41–66. DOI: 10.1111/1475-679X.00038

Breiman, L. (1996a). Bagging predictors. *Machine Learning*, *24*(2), 123–140. DOI: 10.1007/BF00058655

Breiman, L. (1996b). *Stacked Regressions | SpringerLink*. Hentet 20. april 2021, fra https://link.springer.com/article/10.1007/BF00117832

Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. (1984). *Classification And Regression Trees* (1. udg.). Routledge. DOI: 10.1201/9781315139470

Brien, P. C. O., McNichols, M. F. & Lin, H.-W. (2005). Analyst Impartiality and Investment Banking Relationships. *Journal of Accounting Research*, *43*(4), 623–650.

Brown, L. D. (1993). Earnings Forecasting Research: Its Implications for Capital Markets Research. *International Journal of Forecasting*, *9*(3), 295–320. Hentet 7. april 2021, fra https://econpapers.repec.org/article/eeeintfor/v_3a9_3ay_3a1993_3ai_3a3_3ap_3a295-320.htm

Brown, L. D., Call, A. C., Clement, M. B. & Sharp, N. Y. (2014). Inside the 'Black Box' of Sell-Side Financial Analysts. DOI: 10.2139/ssrn.2228373

Brown, L. D., Hagerman, R. L., Griffin, P. A. & Zmijewski, M. E. (1987a). An Evaluation of Alternative Proxies for the Market's Assessment of Unexpected Earnings. *Journal of Accounting and Economics*, *9*(2), 159–193. https://EconPapers.repec.org/RePEc:eee:jaecon:v:9:y:1987:i:2:p:159-193

Brown, L. D., Hagerman, R. L., Griffin, P. A. & Zmijewski, M. E. (1987b). Security analyst superiority relative to univariate time-series models in forecasting quarterly earnings. *Journal of Accounting and Economics*, *9*(1), 61–87. DOI: 10.1016/0165-4101(87)90017-6

Brown, L. D. & Hugon, A. (2009). Team earnings forecasting. *Review of Accounting Studies*, *14*(4), 587–607. DOI: 10.1007/s11142-008-9076-1

Brown, L. D., Richardson, G. D. & Schwager, S. J. (1987). An Information Interpretation of Financial Analyst Superiority in Forecasting Earnings. *Journal of Accounting Research*, *25*(1), 49–67. DOI: 10.2307/2491258

Brown, L. D. & Rozeff, M. S. (1978). The Superiority of Analyst Forecasts as Measures of Expectations: Evidence from Earnings. *Journal of Finance*, *33*(1), 1–16. Hentet 9. februar 2021, fra https://econpapers.repec.org/article/blajfinan/v_3a33_3ay_3a1978_3ai_3a1_3ap_3a1-16.htm

Brown, P. & Kennelly, J. W. (1972). The Informational Content of Quarterly Earnings: An Extension and Some Further Evidence. *The Journal of Business*, *45*(3), 403–415.

Bybee, L., Kelly, B., Manela, A. & Xiu, D. (2020 januar). *The Structure of Economic News* (w26648). National Bureau of Economic Research. Cambridge, MA. DOI: 10.3386/w26648

Call, A. C., Hewitt, M., Watkins, J. & Yohn, T. L. (2016). Changes in the I/B/E/S Database and Their Effect on the Observed Properties of Analyst Forecasts. *SSRN Electronic Journal*. DOI: 10.2139/ssrn.2788140

Capstaff, J., Paudyal, K. & Rees, W. (1995). The Accuracy and Rationality of Earnings Forecasts by Uk Analysts. *Journal of Business Finance & Accounting*, *22*(1), 67–85. DOI: 10.1111/j.1468-5957.1995.tb00672.x
_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-5957.1995.tb00672.x

Capstaff, J., Paudyal, K. & Rees, W. (2001). A Comparative Analysis of Earnings Forecasts in Europe. *Journal of Business Finance & Accounting*, *28*(5-6), 531–562. DOI: 10.1111/1468-5957.00384
_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1468-5957.00384

Capstaff, J. & Rees, K., William Paudyal. (1998). Analysts' Forecasts of German Firms' Earnings: A Comparative Analysis. *Journal of International Financial Management & Accounting*, *9*(2), 83–116. DOI: 10.1111/1467-646X.00032
_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-646X.00032

Carleton, W. T., Chen, C. R. & Steiner, T. L. (1998). Optimism Biases among Brokerage and Non-Brokerage Firms' Equity Recommendations: Agency Costs in the Investment Industry. *Financial Management*, *27*(1), 17–30. DOI: 10.2307/3666148

Casalicchio, G., Molnar, C. & Bischl, B. (2019). *Visualizing the Feature Importance for Black Box Models*. arXiv: 1804.06620 [cs, stat]. DOI: 10.1007/978-3-030-10925-7_40

CatBoost. (2021). *Parameter tuning - CatBoost. Documentation*. Hentet 28. november 2020, fra https://catboost.ai/docs/concepts/parameter-tuning.html

Cavaglia, S., Brightman, C. & Aked, M. (2000). The Increasing Importance of Industry Factors. *Financial Analysts Journal*, *56*, 41–54. DOI: 10.2469/faj.v56.n5.2389

Cecchetti, S. G. & Kashyap, A. K. (1996). International cycles. *European Economic Review*, *40*(2), 331–360. DOI: 10.1016/0014-2921(95)00125-5

CFI Institute. (2021). *Time Period Bias - Overview, Examples, How to Prevent*. Hentet 3. marts 2021, fra https://corporatefinanceinstitute.com/resources/knowledge/other/time-period-bias/

Chopra, V. K. (1998). Why so Much Error in Analysts' Earnings Forecasts? *Financial Analysts Journal*, *54*(6), 35–42.

Chu, X., Ilyas, I. F., Krishnan, S. & Wang, J. (2016). Data Cleaning: Overview and Emerging Challenges. *Proceedings of the 2016 International Conference on Management of Data*, 2201–2206. DOI: 10.1145/2882903.2912574

Claesen, M. & De Moor, B. (2015). *Hyperparameter Search in Machine Learning*. arXiv: 1502.02127 [cs, stat]. Hentet 24. marts 2021, fra http://arxiv.org/abs/1502.02127

Clement, M. B. (1999). Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter? *Journal of Accounting and Economics*, *27*(3), 285–303. DOI: 10.1016/S0165-4101(99)00013-0

Clements, M. P., Galvão, A. B. & Kim, J. H. (2008). Quantile forecasts of daily exchange rate returns from forecasts of realized volatility. *Journal of Empirical Finance*, *15*(4), 729–750. DOI: 10.1016/j.jempfin.2007.12.001

Collins, D. W., Kothari, S. & Rayburn, J. D. (1987). Firm size and the information content of prices with respect to earnings. *Journal of Accounting and Economics*, *9*(2), 111–138. DOI: 10.1016/0165-4101(87)90002-4

Collins, W. A. & Hopwood, W. S. (1980). A Multivariate Analysis of Annual Earnings Forecasts Generated from Quarterly Forecasts of Financial Analysts and Univariate Time-Series Models. *Journal of Accounting Research*, *18*(2), 390–406. DOI: 10.2307/2490585

Craja, P., Kim, A. & Lessmann, S. (2020). Deep learning for detecting financial statement fraud. *Decision Support Systems*, *139*, 113421. DOI: 10.1016/j.dss.2020.113421

CRSP. (2021a januar). *CRSP Indexes - Methodology Guide*.

CRSP. (2021b, 2. februar). *CRSP Monthly Stock*. Hentet 3. marts 2021, fra https://wrds-web.wharton.upenn.edu/wrds//ds/crsp/stock_a/msf.cfm

CRSP. (2021c). *CRSP US Stock Databases | CRSP - The Center for Research in Security Prices*. Hentet 3. marts 2021, fra http://www.crsp.org/products/research-products/crsp-us-stock-databases

CRSP. (2021d). *Data Definitions -CRSP*. Hentet 8. marts 2021, fra http://www.crsp.org/products/documentation/data-definitions-p#price-alternate

CRSP. (2021e). *Why CRSP? | CRSP - The Center for Research in Security Prices*. Hentet 3. marts 2021, fra http://www.crsp.org/main-menu/why-crsp

CRSP & S&P Capital IQ. (2021, 2. marts). *CRSP/Compustat Merged Database - Fundamentals Quarterly*. Hentet 3. marts 2021, fra https://wrds-web.wharton.upenn.edu/wrds//ds/crsp/ccm_a/fundq/index.cfm

DataRobot. (2021). *Automated Machine Learning*. Hentet 30. april 2021, fra https://www.datarobot.com/platform/automated-machine-learning/

De Bondt, W. F. M. & Forbes*, W. P. (1999). Herding in Analyst Earnings Forecasts: Evidence from the United Kingdom. *European Financial Management*, *5*(2), 143–163. Hentet 25. januar 2021, fra https://econpapers.repec.org/article/blaeufman/v_3a5_3ay_3a1999_3ai_3a2_3ap_3a143-163.htm

Dechow, P. M., Hutton, A. P. & Sloan, R. G. (2000). The Relation between Analysts' Forecasts of Long-Term Earnings Growth and Stock Price Performance Following Equity Offerings*. *Contemporary Accounting Research*, *17*(1), 1–32. DOI: 10.1111/j.1911-3846.2000.tb00908.x _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1911-3846.2000.tb00908.x

Devos, E. (2014). Are Analysts' Recommendations for Other Investment Banks Biased? *Financial Management*, *43*(2), 327–353. DOI: 10.1111/fima.12031 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/fima.12031

Dhaliwal, D. S., Radhakrishnan, S., Tsang, A. & Yang, Y. G. (2012). Nonfinancial Disclosure and Analyst Forecast Accuracy: International Evidence on Corporate Social Responsibility Disclosure. *The Accounting Review*, *87*(3), 723–759.

Diebold, F. & Mariano, R. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–63. Hentet 1. maj 2021, fra https://econpapers.repec.org/article/besjnlbes/v_3a13_3ay_3a1995_3ai_3a3_3ap_3a253-63.htm

Dietterich, T. G. (2000). An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization. *Machine Learning*, *40*(2), 139–157. DOI: 10.1023/A:1007607513941

Dixon, M., Klabjan, D. & Bang, J. H. (2017). Classification-Based Financial Markets Prediction Using Deep Neural Networks. *Algorithmic Finance*, *6*(3-4), 67–77. DOI: 10.3233/AF-170176

D'Onfro, J. (2019). *AI 50: America's Most Promising Artificial Intelligence Companies*. Hentet 17. marts 2021, fra https://www.forbes.com/sites/jilliandonfro/2019/09/17/ai-50-americas-most-promising-artificial-intelligence-companies/?sh=22574ba0565c

Dong, Y. & Peng, C.-Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus*, *2*(1), 222. DOI: 10.1186/2193-1801-2-222

Dorogush, A. V., Ershov, V. & Gulin, A. (2018, 24. oktober). *CatBoost: Gradient Boosting with Categorical Features Support*. arXiv: 1810.11363 [cs, stat]. Hentet 27. marts 2021, fra http://arxiv.org/abs/1810.11363

Drechsler, J. (2020). Multiple Imputation. *SAGE Research Methods Foundations*. SAGE Publications Ltd. DOI: 10.4135/9781526421036885886

Duarte, C., Rodrigues, P. M. & Rua, A. (2017). A mixed frequency approach to the forecasting of private consumption with ATM/POS data. *International Journal of Forecasting*, *33*(1), 61–75. DOI: 10.1016/j.ijforecast.2016.08.003

Elend, L., Tideman, S. A., Lopatta, K. & Kramer, O. (2020). Earnings Prediction with Deep Leaning. I U. Schmid, F. Klügl & D. Wolter (Red.), *KI 2020: Advances in Artificial Intelligence* (s. 267–274). Springer International Publishing. DOI: 10.1007/978-3-030-58285-2_22

Ershov, V. (2018). *CatBoost Enables Fast Gradient Boosting on Decision Trees Using GPUs*. Hentet 27. marts 2021, fra https://catboost.ai/news/catboost-enables-fast-gradient-boosting-on-decision-trees-using-gpus

Faggella, D. (2020, 3. april). *Machine Learning in Finance - Present and Future Applications*. Hentet 10. maj 2021, fra https://emerj.com/ai-sector-overviews/machine-learning-in-finance/

Fama, E. & French, K. R. (2021). *Kenneth R. French - Detail for 12 Industry Portfolios*. Hentet 8. marts 2021, fra http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_12_ind_port.html

Fang, C. & Wang, C. (2020, 23. november). *Time Series Data Imputation: A Survey on Deep Learning Approaches*. arXiv: 2011.11347 [cs]. Hentet 8. marts 2021, fra http://arxiv.org/abs/2011.11347

Foster, G. (1977). Quarterly Accounting Data: Time-Series Properties and Predictive-Ability Results. *The Accounting Review*, *52*(1), 1–21.

Freeman, R. N. (1987). The association between accounting earnings and security returns for large and small firms. *Journal of Accounting and Economics*, *9*(2), 195–228. DOI: 10.1016/0165-4101(87)90005-X

Fried, D. & Givoly, D. (1982). Financial analysts' forecasts of earnings: A better surrogate for market expectations. *Journal of Accounting and Economics*, *4*(2), 85–107. DOI: 10.1016/0165-4101(82)90015-5

Galanti, S. & Vaubourg, A. G. (2017). Optimism bias in financial analysts' earnings forecasts: Do commissions sharing agreements reduce conflicts of interest? *Economic Modelling*, *67*, 325–337. DOI: 10.1016/j.econmod.2017.02.001

Geremew, M. & Gourio, F. (2018). Seasonal and Business Cycles of U.S. Employment. *Economic Perspectives*. DOI: 10.21033/ep-2018-3

Ghosh, D. & Vogt, A. (2012). Outliers: An Evaluation of Methodologies, 6.

Ghysels, E., Santa-Clara, P. & Valkanov, R. (2004, 1. maj). *The MIDAS Touch: Mixed Data Sampling Regression Models* (2004s-20). CIRANO. Hentet 10. februar 2021, fra https://ideas.repec.org/p/cir/cirwor/2004s-20.html

Ghysels, E., Santa-Clara, P. & Valkanov, R. (2005). There is a risk-return trade-off after all. *Journal of Financial Economics*, *76*(3), 509–548. DOI: 10.1016/j.jfineco.2004.03.008

Givoly, D. & Lakonishok, J. (1979). The information content of financial analysts' forecasts of earnings: Some evidence on semi-strong inefficiency. *Journal of Accounting and Economics*, *1*(3), 165–185. DOI: 10.1016/0165-4101(79)90006-5

Graham, J. R. (nodate). Herding among Investment Newsletters: Theory and Evidence. *The Journal of Finance*, 33.

Gunay, S., Can, G. & Ocak, M. (2020). Forecast of China's economic growth during the COVID-19 pandemic: A MIDAS regression analysis. *Journal of Chinese Economic and Foreign Trade Studies*, *ahead-of-print*. DOI: 10.1108/JCEFTS-08-2020-0053

Gupta, P. (2017, 5. juni). *Cross-Validation in Machine Learning*. Hentet 24. marts 2021, fra https://towardsdatascience.com/cross-validation-in-machine-learning-72924a69872f

Hancock, J. T. & Khoshgoftaar, T. M. (2020). CatBoost for Big Data: An Interdisciplinary Review. *Journal of Big Data*, *7*(1), 94. DOI: 10.1186/s40537-020-00369-8

Harvey, D., Leybourne, S. & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, *13*(2), 281–291. DOI: 10.1016/S0169-2070(96)00719-4

Hastie, T., Tibshirani, R. & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition* (2. udg.). Springer-Verlag. DOI: 10.1007/978-0-387-84858-7

Holland, J. B. (1998). Private disclosure and financial reporting. *Accounting and Business Research*, *28*(4), 255–269. DOI: 10.1080/00014788.1998.9728914

Holthausen, R. W. & Larcker, D. F. (1992). The Prediction of Stock Returns Using Financial Statement Information. *Journal of Accounting and Economics*, *15*(2-3), 373–411. https://EconPapers.repec.org/RePEc:eee:jaecon:v:15:y:1992:i:2-3:p:373-411

Honaker, J. & King, G. (2010). What to Do about Missing Values in Time-Series Cross-Section Data. *American Journal of Political Science*, *54*(2), 561–581. DOI: 10.1111/j.1540-5907.2010.00447.x

Hong, H., Kubik, J. D. & Solomon, A. (2000). Security Analysts' Career Concerns and Herding of Earnings Forecasts. *The RAND Journal of Economics*, *31*(1), 121–144. DOI: 10.2307/2601032

Hope, O.-K. (2003). Disclosure Practices, Enforcement of Accounting Standards, and Analysts' Forecast Accuracy: An International Study. *Journal of Accounting Research*, *41*(2), 235–272. DOI: 10.1111/1475-679X.00102

Hope, O.-K. & Kang, T. (2005). The Association between Macroeconomic Uncertainty and Analysts' Forecast Accuracy. *Journal of International Accounting Research - J Int Account Res*, *4*, 23–38. DOI: 10.2308/jiar.2005.4.1.23

Hou, K., van Dijk, M. A. & Zhang, Y. (2012). The implied cost of capital: A new approach. *Journal of Accounting and Economics*, *53*(3), 504–526. DOI: 10.1016/j.jacceco.2011.12.001

Hou, T. C.-T., Hung, W. & Gao, S. S. (2014). Investors' Reactions to Analysts' Forecast Revisions and Information Uncertainty: Evidence of Stock Price Drift. *Journal of Accounting, Auditing & Finance*, *29*(3), 238–259. DOI: 10.1177/0148558X14530129

Huberts, L. C. & Fuller, R. J. (1995). Predictability Bias in the U.S. Equity Market. *Financial Analysts Journal*, *51*(2), 12–28.

Hull, J. (2020). Machine Learning and Finance. *Journal of Risk Management in Financial Institutions*, *13*(2), 104–105. http://esc-web.lib.cbs.dk/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=142397566&site=ehost-live&scope=site

Hunt, J., Myers, J. & Myers, L. (2019). Improving Earnings Predictions with Machine Learning. Hentet 14. februar 2021, fra https://zicklin.baruch.cuny.edu/wp-content/uploads/sites/10/2019/12/Improving-Earnings-Predictions-with-Machine-Learning-Hunt-Myers-Myers.pdf

Hyndman, R. J. & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd). OTexts. Hentet 16. marts 2021, fra https://otexts.com/fpp3/

Hyndman, R. J. & Khandakar, Y. (2008). Automatic Time Series Forecasting: The Forecast Package for R. *Journal of Statistical Software*, *26*(3), 1–22. https://www.jstatsoft.org/article/view/v027i03

Hyndman, R. J. & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688. DOI: 10.1016/j.ijforecast.2006.03.001

IBES & Refinitiv. (2021). *IBES - Eikon - Python API*. Hentet 3. marts 2021, fra https://developers.refinitiv.com/en/api-catalog/eikon/eikon-data-api

J. Armstrong, S. (2001). *Principles of Forecasting A Handbook for Researchers and Practitioners*. Hentet 16. marts 2021, fra https://doi.org/10.1007/978-0-306-47630-3
OCLC: 1075139423

Jackson, A. R. (2005). Trade Generation, Reputation, and Sell-Side Analysts. *The Journal of Finance*, *60*(2), 673–717.

Jacob, J. (1997). Discussion of Do Security Analysts Improve Their Performance with Experience? *Journal of Accounting Research*, *35*, 159. DOI: 10.2307/2491459

Jacob, J., Lys, T. Z. & Neale, M. A. (1999). Expertise in forecasting performance of security analysts. *Journal of Accounting and Economics*, *28*(1), 51–82. DOI: 10.1016/S0165-4101(99)00016-6

Jacob, J., Rock, S. & Weber, D. P. (2008). Do Non-Investment Bank Analysts Make Better Earnings Forecasts? *Journal of Accounting, Auditing & Finance*, *23*(1), 23–61. DOI: 10.1177/0148558X0802300104

John, C., Ekpenyong, E. J. & Nworu, C. C. (2019). Imputation of Missing Values in Economic and Financial Time Series Data Using Five Principal Component Analysis (PCA) Approaches.

Jung, M. J., Keeley, J. H. & Ronen, J. (2019). The Predictability of Analyst Forecast Revisions. *Journal of Accounting, Auditing & Finance*, *34*(3), 434–457. DOI: 10.1177/0148558X17722710

Kaplan, Z., Martin, X. & Xie, Y. (2018). Truncating Optimism. *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3213833

Kenton, W. (2019, 11. september). *Standard Industrial Classification (SIC Code)*. Hentet 9. marts 2021, fra https://www.investopedia.com/terms/s/sic_code.asp

Kerl, A. & Ohlert, M. (2015). STAR-ANALYSTS' FORECAST ACCURACY AND THE ROLE OF CORPORATE GOVERNANCE: Star-Analysts' Forecast Accuracy. *Journal of Financial Research*, *38*(1), 93–120. DOI: 10.1111/jfir.12053

Kerl, A. G. & Ohlert, M. (2014, 28. august). *Star-Analysts' Forecast Accuracy and the Role of Corporate Governance* (SSRN Scholarly Paper Nr. ID 2195909). Social Science Research Network. Rochester, NY. DOI: 10.2139/ssrn.2195909

Keskek, S., Tse, S. & Tucker, J. W. (2014). Analyst information production and the timing of annual earnings forecasts. *Review of Accounting Studies*, *19*(4), 1504–1531. DOI: 10.1007/s11142-014-9278-7

Keynes, J. M. (1936). *The General Theory of Employment, Interest, and Money*.

Koonce, L. (1997). Discussion of Implications of Proposed Segment Reporting Standards for Financial Analysts' Investment Judgments. *Journal of Accounting Research*, *35*, 25–33. DOI: 10.2307/2491451

Kothari, S. P. (2001). Capital markets research in accounting. *Journal of Accounting and Economics*, *31*(1), 105–231. DOI: 10.1016/S0165-4101(01)00030-1

Krane, S. & Wascher, W. (1999). The cyclical sensitivity of seasonality in U.S. employment. *Journal of Monetary Economics*, *44*(3), 523–553. DOI: 10.1016/S0304-3932(99)00036-7

Kraus, M. & Feuerriegel, S. (2017). Decision Support from Financial Disclosures with Deep Neural Networks and Transfer Learning. *Decision Support Systems*, *104*, 38–48. DOI: 10.1016/j.dss.2017.10.001

Kross, W., Ro, B. & Schroeder, D. (1990). Earnings Expectations: The Analysts' Information Advantage. *The Accounting Review*, *65*(2), 461–476.

Kuhn, M. & Johnson, K. (2013). *Applied Predictive Modeling*. Springer
OCLC: ocn827083441.

Kwag, S.-W. (& Small, K. (2007). The impact of regulation fair disclosure on earnings management and analyst forecast bias. *Journal of Economics and Finance*, *31*(1), 87–98. DOI: 10.1007/BF02751514

Lang, M. H. & Lundholm, R. J. (1996). Corporate Disclosure Policy and Analyst Behavior. *The Accounting Review*, *71*(4), 467–492.

Lang, M. H., Pinto, J. & Sul, E. (2019). MiFID II Unbundling and Sell Side Analyst Research. *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3408198

Leece, R. D. & White, T. P. (2017). The Effects of Firms' Information Environment on Analysts' Herding Behavior. *Review of Quantitative Finance and Accounting*, *48*(2), 503–525. Hentet

26. januar 2021, fra https://econpapers.repec.org/article/kaprqfnac/v_3a48_3ay_3a2017_3ai_3a2_3ad_3a10.1007_5fs11156-016-0559-z.htm

Leung, S. & Srinidhi, B. (2006). The Effect of the Private Securities Litigation Reform Act on Analyst Forecast Properties: The Impact of Firm Size and Growth Opportunities. *Journal of Business Finance & Accounting*, *33*(5-6), 767–792. DOI: 10.1111/j.1468-5957.2006.00020.x _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-5957.2006.00020.x

Lev, B. & Thiagarajan, S. R. (23–1993). Fundamental Information Analysis. *Journal of Accounting Research*, *31*(2), 190. DOI: 10.2307/2491270

Li, K. K. & Mohanram, P. (2014). Evaluating cross-sectional forecasting models for implied cost of capital. *Review of Accounting Studies*, *19*(3), 1152–1185. DOI: 10.1007/s11142-014-9282-y

Lin, W.-C. & Tsai, C.-F. (2020). Missing value imputation: A review and analysis of the literature (2006–2017). *Artificial Intelligence Review*, *53*(2), 1487–1509. DOI: 10.1007/s10462-019-09709-4

Little, I. M. D. (1962). Higgledy Piggledy Growth. *Bulletin of the Oxford University Institute of Economics & Statistics*, *24*(4), 387–412. DOI: 10.1111/j.1468-0084.1962.mp24004001.x _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0084.1962.mp24004001.x

Liu, C., Gao, C., Xia, X., Lo, D., Grundy, J. & Yang, X. (2020). On the Replicability and Reproducibility of Deep Learning in Software Engineering. *ArXiv*.

Ljungqvist, A., Malloy, C. & Marston, F. (2009). Rewriting History. *The Journal of Finance*, *64*(4), 1935–1960. DOI: 10.1111/j.1540-6261.2009.01484.x

Lorek, K. S. (1979). Predicting Annual Net Earnings with Quarterly Earnings Time-Series Models. *Journal of Accounting Research*, *17*(1), 190–204. DOI: 10.2307/2490313

Lundberg, S. M. & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. I I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan & R. Garnett (Red.), *Advances in Neural Information Processing Systems 30* (s. 4765–4774). Curran Associates, Inc. http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

Lys, T. & Soo, L. G. (1995). Analysts' Forecast Precision as a Response to Competition. *Journal of Accounting, Auditing & Finance*, *10*(4), 751–765. DOI: 10.1177/0148558X9501000404

M. Madhani, P. (2016). *Firm Size, Corporate Governance and Disclosure Practices: Inter-Relations*. Hentet 13. april 2021, fra https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2803036

Maines, L. A., McDaniel, L. S. & Harris, M. S. (1997). Implications of Proposed Segment Reporting Standards for Financial Analysts' Investment Judgements. *Journal of Accounting Research*, *35*, 1. DOI: 10.2307/2491450

Matas-Mir, A. & Osborn, D. R. (2004). Does seasonality change over the business cycle? An investigation using monthly industrial production series. *European Economic Review*, *48*(6), 1309–1332. DOI: 10.1016/j.euroecorev.2003.10.003

McKinsey Global Institute. (2018 september). *NOTES FROM THE AI FRONTIER MODELING THE IMPACT OF AI ON THE WORLD ECONOMY*. McKinsey & Company. Hentet 8. maj 2021, fra https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Notes%20from%20the%20frontier%20Modeling%20the%20impact%20of%20AI%20on%20the%20world%20economy/MGI-Notes-from-the-AI-frontier-Modeling-the-impact-of-AI-on-the-world-economy-September-2018.ashx

McKnight, P. J., Tavakoli, M. & Weir, C. (2010). Underwriting Relationships and Analyst Independence in Europe. DOI: 10.1111/j.1468-0416.2010.00157.x

Mehran, H. & Stulz, R. M. (2007). The economics of conflicts of interest in financial institutions. *Journal of Financial Economics*, *85*(2), 267–296. DOI: 10.1016/j.jfineco.2006.11.001

Michaely, R. & Womack, K. L. (1999). Conflict of Interest and the Credibility of Underwriter Analyst Recommendations. *The Review of Financial Studies*, *12*(4), 653–686. DOI: 10.1093/rfs/12.4.653

Michel, J.-S. (2017). Investor Overreaction to Analyst Reference Points. *Journal of Behavioral Finance*, *18*(3), 329–343. DOI: 10.1080/15427560.2017.1342646

Mikhail, M. B., Walther, B. R. & Willis, R. H. (1997). Do Security Analysts Improve Their Performance with Experience? *Journal of Accounting Research*, *35*, 131–157. DOI: 10.2307/2491458

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.

Müller, A. C. & Guido, S. (2016). *Introduction to Machine Learning with Python: A Guide for Data Scientists*. O'Reilly Media, Incorporated. Hentet 19. april 2021, fra http://ebookcentral.proquest.com/lib/kbhnhh-ebooks/detail.action?docID=4698164

Myring, M. & Wrege, W. (2009). Analysts Earnings Forecast Accuracy And Activity: A Time-Series Analysis. *Journal of Business & Economics Research (JBER)*, *7*(5). DOI: 10.19030/jber.v7i5.2295

Nau, R. (2020, 18. august). *Statistical Forecasting: Notes on Regression and Time Series Analysis*. Hentet 10. februar 2021, fra https://people.duke.edu/~rnau/411home.htm

N.Dreman, D. & Berry, M. A. (1995). Analyst Forecasting Errors and Their Implications for Security Analysis. *Financial Analysts Journal*, *51*(3), 30–41.

Newey, W. K. & West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, *55*(3), 703. DOI: 10.2307/1913610

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, *108*(1), 1–28. DOI: 10.1016/j.jfineco.2013.01.003

O'brien, P. C. (1988). Analysts' forecasts as earnings expectations. *Journal of Accounting and Economics*, *10*(1), 53–83. DOI: 10.1016/0165-4101(88)90023-7

O'REILLY. (2021). *4. Hyperparameter Tuning - Evaluating Machine Learning Models [Book]*. Hentet 1. maj 2021, fra https://www.oreilly.com/library/view/evaluating-machine-learning/9781492048756/ch04.html

Ou, J. A. & Penman, S. H. (1989). Financial statement analysis and the prediction of stock returns. *Journal of Accounting and Economics*, *11*(4), 295–329. DOI: 10.1016/0165-4101(89)90017-7

Panda, M. & Misra, H. (Red.). (2021). *Handbook of Research on Automated Feature Engineering and Advanced Applications in Data Science:* IGI Global. DOI: 10.4018/978-1-7998-6659-6

Payne, J. L. & Thomas, W. B. (2003). The Implications of Using Stock-Split Adjusted I/B/E/S Data in Empirical Research. *The Accounting review*, *78*(4), 1049–1067. DOI: 10.2308/accr.2003.78.4.1049

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V. & Gulin, A. (2017). *CatBoost: Unbiased Boosting with Categorical Features*. arXiv: 1706.09516 [cs]. Hentet 27. marts 2021, fra http://arxiv.org/abs/1706.09516

Rajan, R. G. (2006). Has Finance Made the World Riskier? *European Financial Management*, *12*(4), 499–533. DOI: 10.1111/j.1468-036X.2006.00330.x

Reback, J., McKinney, W., Jbrockmendel, Bossche, J. V. D., Augspurger, T., Cloud, P., Gfyoung, Sinhrks, Klein, A., Roeschke, M., Hawkins, S., Tratner, J., She, C., Ayd, W., Petersen, T.,

Garcia, M., Schendel, J., Hayden, A., MomIsBestFriend, … Mehyar, M. (2021, 2. marts). *Pandas - Version 1.2.3* (Vers. v1.2.3). DOI: 10.5281/ZENODO.3715232

Refinitiv. (nodate). *Institutional Brokers Estimate System IBES*. Hentet 25. februar 2021, fra https://www.refinitiv.com/en/financial-data/company-data/institutional-brokers-estimate-system-ibes

Refinitiv. (2021). *StarMine SmartEstimates*. Hentet 26. april 2021, fra https://www.refinitiv.com/en/financial-data/analytics/quantitative-analytics/starmine-smartestimates

Ruder, S., Peters, M. E., Swayamdipta, S. & Wolf, T. (2019). Transfer Learning in Natural Language Processing. *Proceedings of the 2019 Conference of the North*, 15–18. DOI: 10.18653/v1/N19-5004

Run:AI. (2021). *Deep Learning with GPUs*. Hentet 9. marts 2021, fra https://www.run.ai/guides/gpu-deep-learning/

Scharfstein, D. & Stein, J. (1990). Herd Behavior and Investment. *American Economic Review*, *80*(3), 465–79. Hentet 26. januar 2021, fra https://econpapers.repec.org/article/aeaaecrev/v_3a80_3ay_3a1990_3ai_3a3_3ap_3a465-79.htm

Schreiner, A. & Spremann, K. (2007). Multiples and Their Valuation Accuracy in European Equity Markets. *SSRN Electronic Journal*. DOI: 10.2139/ssrn.957352

Scikit Learn. (2020). *Sklearn.Model_selection.TimeSeriesSplit — Scikit-Learn 0.24.2 Documentation*. Hentet 30. april 2021, fra https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.TimeSeriesSplit.html

Scikit-learn. (2020a). *Linear Models*. Hentet 21. april 2021, fra https://scikit-learn.org/stable/modules/svm.html#svm-regression

Scikit-learn. (2020b). *Support Vector Machines*. Hentet 21. april 2021, fra https://scikit-learn.org/stable/modules/svm.html#svm-regression

SEC. (2003, 28. april). *SEC vs Goldman Sachs - Complaint*. Hentet 26. januar 2021, fra https://www.sec.gov/litigation/complaints/comp18113.htm

Seng, D. & Hancock, J. R. (2012). Fundamental Analysis and the Prediction of Earnings. *International Journal of Business and Management*, *7*(3), p32. DOI: 10.5539/ijbm.v7n3p32

Sermpinis, G., Laws, J., Karathanasopoulos, A. & Dunis, C. L. (2012). Forecasting and trading the EUR/USD exchange rate with Gene Expression and Psi Sigma Neural Networks. *Expert Systems with Applications*, *39*(10), 8865–8877. DOI: 10.1016/j.eswa.2012.02.022

Sezer, O. B., Gudelek, M. U. & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning : A systematic literature review: 2005–2019. *Applied Soft Computing*, *90*, 106181. DOI: 10.1016/j.asoc.2020.106181

Sharot, T. (2011). The optimism bias. *Current Biology*, *21*(23), R941–R945. DOI: 10.1016/j.cub.2011.10.030

Shekhar, A. (2018). *What Is Feature Engineering for Machine Learning?* Hentet 22. marts 2021, fra https://medium.com/mindorks/what-is-feature-engineering-for-machine-learning-d8ba3158d97a

Shen, J. & Shafiq, M. O. (2020). Short-term stock market price trend prediction using a comprehensive deep learning system. *Journal of Big Data*, *7*(1), 66. DOI: 10.1186/s40537-020-00333-6

Sonne, E. (2021, 9. februar). *LibGuides: WRDS: WRDS - Content*. Hentet 3. marts 2021, fra https://libguides.cbs.dk/c.php?g=417474&p=2845102

Stickel, S. E. (23–1990). Predicting Individual Analyst Earnings Forecasts. *Journal of Accounting Research*, *28*(2), 409. DOI: 10.2307/2491158

Stickel, S. E. (1992). Reputation and Performance Among Security Analysts. *The Journal of Finance*, *47*(5), 1811–1836. DOI: 10.1111/j.1540-6261.1992.tb04684.x _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1992.tb04684.x

Subrahmanyam, A. (2019). Big data in finance: Evidence and challenges. *Borsa Istanbul Review*, *19*(4), 283–287. DOI: 10.1016/j.bir.2019.07.007

Tableau Software LLC. (2021). *Data cleaning: The benefits and steps to creating and using clean data*. Hentet 9. marts 2021, fra https://www.tableau.com/learn/articles/what-is-data-cleaning

Tibshirani, S. & Friedman, H. (nodate). Valerie and Patrick Hastie, 764.

Tong, H. (2007). Disclosure standards and market efficiency: Evidence from analysts' forecasts. *Journal of International Economics*, *72*(1), 222–241. DOI: 10.1016/j.jinteco.2006.11.004

Trueman, B. (1994). Analyst Forecasts and Herding Behavior. *Review of Financial Studies*, *7*(1), 97–124. Hentet 26. januar 2021, fra https : //econpapers.repec.org/article/ouprfinst/v_3a7_3ay_3a1994_3ai_3a1_3ap_3a97-124.htm

TUKEY, J. W. (1960). A Survey of Sampling from Contaminated Distributions. *Contributions to Probability and Statistics*, 448–485. https://ci.nii.ac.jp/naid/20000755025/en/

U.S securities and exchange commision. (2021). *Form 10-Q | Investor.Gov*. Hentet 26. april 2021, fra https://www.investor.gov/introduction-investing/investing-basics/glossary/form-10-q

van Binsbergen, J. H., Han, X. & Lopez Lira, A. (2020). Man versus Machine Learning: Earnings Expectations and Conditional Biases. *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3625279

Vargas, M. R., de Lima, B. S. L. P. & Evsukoff, A. G. (2017). Deep Learning for Stock Market Prediction from Financial News Articles. *2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, 60–65. DOI: 10.1109/CIVEMSA.2017.7995302

Welch, I. (2000). Herding among Security Analysts. *Journal of Financial Economics*, *58*(3), 369–396. Hentet 26. januar 2021, fra https://econpapers.repec.org/article/eeejfinec/v_3a58_3ay_3a2000_3ai_3a3_3ap_3a369-396.htm

Wolpert, D. (1992). Stacked Generalization. *Neural Networks*, *5*, 241–259. DOI: 10.1016/S0893-6080(05)80023-1

WRDS. (2021). *WRDS Overview of Compustat North America - Global - Bank*. Hentet 3. marts 2021, fra https://wrds-www.wharton.upenn.edu/pages/support/manuals-and-overviews/compustat/north-america-global-bank/wrds-overview-compustat-north-america-global-bank/

Wright, J. H. (2013). Unseasonal Seasonals? *Brookings Papers on Economic Activity*, *2013*(2), 65–126. DOI: 10.1353/eca.2013.0017

Xie, Y., Zhang, Y. & Zhou, J. (2012). National Level, City Level Auditor Industry Specialization and Analyst Forecast Properties: National Level, City Level Auditor Industry Specialization and Analyst Forecast Properties. *International Journal of Auditing*, *16*(3), 248–267. DOI: 10.1111/j.1099-1123.2012.00449.x

Xinyue, C., Zhaoyu, X. & Yue, Z. (2021). Using Machine Learning to Forecast Future Earnings. *Atlantic Economic Journal*. DOI: 10.1007/s11293-020-09691-1

Yan, X. (& Zheng, L. (2017). Fundamental Analysis and the Cross-Section of Stock Returns: A Data-Mining Approach. *The Review of Financial Studies*, *30*(4), 1382–1423. DOI: 10.1093/rfs/hhx001

Yu, M. (2010). Analyst forecast properties, analyst following and governance disclosures: A global perspective. *Journal of International Accounting, Auditing and Taxation*, *19*(1), 1–15. DOI: 10.1016/j.intaccaudtax.2009.12.001

Zhang, Y.-J. & Wang, J.-L. (2019). Do high-frequency stock market data help forecast crude oil prices? Evidence from the MIDAS models. *Energy Economics*, *78*, 192–201. DOI: 10.1016/j.eneco.2018.11.015

Zheng, A. & Casari, A. (2018). *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists* (First edition). O'Reilly
OCLC: ocn957747646.