

MASTEROPPGAVE

Emnekode: BE320E

Kandidatnummer 106 og 110: Truls Didriksen og Truls Rødfjell

Daglig utvikling på Oslo Børs Hovedindeks: En maskinlæringstilnærming

Dato: 07.12.2018

Totalt antall sider: 93

ABSTRACT

In our thesis we have used machine learning to predict day to day change on Oslo Stock Exchange. We've used, among others, daily change in stock indices, oil price and currency as independent variables to try to predict if OSEBX will go up or down by the end of the same day. The first research question is to what extent it is possible to predict the movement on OSEBX based on our selected variables. Our results show that we are able to predict correct result with an accuracy just below 77 %.

Our second research question is which of our selected machine learning methods is most suitable to answer our first research question. We found that, in general, the linear methods performed slightly better than the non-linear methods. The differences are quite marginal, but since linear methods give us an idea of which variables are most important, and they perform marginally better, we found that the best method to solve our problem was logistic regression.

Our third research question is formed to identify which of our independent variables are most important to explain OSEBX. We found that oil price and the stock indices were the most important. Currency, turnover and volatility was not very important.

FORORD

Denne studien er en Masteroppgave som utgjør 30 studiepoeng og er vårt avsluttende arbeid tilknyttet MBA-studiet på Handelshøgskolen ved Nord universitet. Tiden som student ved siden av full jobb har vært en krevende tid, men også veldig lærerik og inspirerende.

Arbeidet med studiet har vært en lærerik tid. Vi har både fått utvidet horisonten innenfor fagfeltene finans og maskinlæring, og vi har lært mye av den prosessen det er å gjennomføre en slik studie.

Vi vil takke vår veileder, førsteamanuensis Thomas Leirvik, for presis veiledning og gode råd underveis i prosessen, og vi vil takke førsteamanuensis Andreas Mikkelsen for å gitt oss tilgang til datamaterialet.

Til slutt vil vi takke hverandre for godt samarbeid gjennom hele studiet, og for å ha holdt humøret uansett hvor lavt blodsukkeret er, eller hvor mange timer vi har jobbet.

Handelshøgskolen ved Nord universitet, 6. desember 2018



Truls Didriksen



Truls Rødfjell

SAMMENDRAG

Målet med dette studiet tredelt. Det første forskningsspørsmålet vårt knytter seg til hvorvidt, og i hvor stor grad, vi klarer å si noe om dag-til-dag utviklingen på Oslo Børs, representert ved hovedindeksen. For å få svar på dette har vi analysert et utvalg variabler ved hjelp av ulike maskinlæringsmetoder.

Det andre forskningsspørsmålet vårt går ut på å finne ut hvilken av våre utvalgte maskinlæringsmetoder som er best egnet til å gi svar på vårt spørsmål. Vi har benyttet åtte ulike maskinlæringsalgoritmer. Algoritmene kan deles i to hovedkategorier; Lineære algoritmer, som gir en indikasjon på hvilke variabler som er de viktigste i analysen, og Ikke-lineære som leverer modeller som er mer komplekse, og derfor vanskeligere å tolke.

Det tredje forskningsspørsmålet vårt er knyttet til de lineære metodene. Vi har forsøkt å gi svar på hvilke av våre utvalgte variabler som er viktigst for å forklare utviklingen på Oslo Børs.

For å gi ett godt grunnlag for å svare på spørsmålene våre gjør vi rede for relevant økonomisk teori og maskinlæringsteori, før vi går igjennom våre valgte variabler. Vi har valgt en bredt utvalg aksjeindekser, oljepris og valutakurser i tillegg til omsetningstall fra Oslo Børs og volatilitetsindeksen VIX. Vi har et datautvalg med daglige verdier i perioden 2001-2017.

I maskinlæring er det vanlig å dele datasettet – en del for å generere modeller, og en del for å teste de genererte modellene mot usett data. I vår studie har vi delt datasettet i to deler ved hjelp av stratifisert deling. Det vil si at datasettet deles i to tilfeldige deler med lik fordeling av det utfallet man ønsker å predikere. Vi valgte å dele datasettet på tre ulike måter, slik at vi fikk testet metodene på flere datautvalg. Dette gjorde vi for å unngå at en uheldig deling påvirker våre funn. I alle tre tilfellene fikk metodene 70 % treningsdata og 30 % testdata. Det vil si at metodene benytter 70 % av datasettet for å generere en modell som skal være velegnet til å predikere både på testdata og andre datasett med samme variabler.

Vår analyse viser at modellene presterer nokså likt når det kommer til antallet riktige prediksjoner. I beste tilfelle klarer modellen å predikere riktig utfall i 77 % av tilfellene, mens den modellen som presterer dårligst klarer å predikere riktig i ca. 74 % av tilfellene.

Maskinlæringsmetodene deles som nevnt i to kategorier, lineære og ikke-lineære metoder. Vi har et metodeutvalg som består av tre lineære og fem ikke-lineære metoder. Våre funn tyder på at de lineære metodene er noe bedre egnet til å gi svar på utviklingen på Oslo Børs, basert på våre variabler. I tillegg til at de scorer bedre på både nøyaktighet (Accuracy) og presisjon

(Precision), så har de også lavere log loss, noe som tyder på at prediksjonene gjøres med høyere kvalitet enn de ikke-lineære. Den metoden vi mener presterte aller best i vår studie var Logistic Regression.

De lineære metodene gir også en indikasjon på hvilke av variablene som er viktigst, ved at variablene får vekt i modellene som genereres. Alle de tre lineære metodene har nokså like resultater tilknyttet hvilke variabler som er viktigst. Vi ser at det er de geografisk nære børsene som vektlegges tyngst, herunder Sverige, Danmark og Storbritannia. I tillegg er USA nokså viktig. Oljepris vektet på linje med de viktigste børsene, mens valutakurs er mindre viktig. Vi har ikke funnet at volatilitet eller omsetning på Oslo Børs har nevneverdig vekt i noen av modellene.

Innhold

1. Innledning	1
1.1. Problemstilling.....	2
1.2. Oppgavens oppbygging	3
2. Teori.....	4
2.1. Økonomisk teori	4
2.1.1. Om aksjemarkedet.....	4
2.1.2. Forventninger og usikkerhet.....	4
2.1.3. Markedseffisiens	5
2.1.4. Kapitalverdimodellen KVM.....	6
2.1.5. Arbitrasjepricingsteorien	6
2.1.6. Random walk.....	7
2.2. Maskinlæring	8
2.2.1. Definisjon	8
2.2.2. Måling av ytelse	9
2.2.3. Varians vs skjevhet.....	11
2.2.4. Utelatt variabel-skjevhet	11
2.2.5. Maskinlæringsalgoritmer	11
3. Variabler	17
3.1. Avhengig variabel.....	17
3.2. Uavhengige variabler.....	18
4. Metode	22
4.1. Datatransformering	23
4.1.1. Åpningstider på børsene.....	23
4.1.2. Oppbygging av datasett.....	24
4.1.3. Editering av Metadata	28
4.1.4. Normalisering av data for lineære modeller.....	29
4.2. Splitting av data	29
4.3. Generering av modell	30
4.4. Evaluering av modell.....	31
4.4.1. Accuracy.....	32
4.4.2. Precision	32
4.4.3. Recall.....	32
4.4.4. F-Score	33
4.4.5. Area under the curve (AUC)	33
4.4.6. Logarithmic Loss.....	35
4.5. Reliabilitet	36

4.6.	Validitet	36
4.6.1.	Intern validitet	36
4.6.2.	Ekstern validitet.....	37
4.7.	Forskningsetikk	37
5.	Resultat og analyse	38
5.1.	Lineære metoder	38
5.1.1.	Bayes Point Machine.....	39
5.1.2.	Logistic Regression	42
5.1.3.	Support vector machine.....	45
5.1.4.	Sammenligning lineære metoder.....	48
5.2.	Ikke-lineære metoder	50
5.2.1.	Locally-Deep Support Vector Machine	50
5.2.2.	Boosted Decision Tree	51
5.2.3.	Decision Forest.....	53
5.2.4.	Decision Jungle	54
5.2.5.	Neural Network	56
5.3.	Analyse av måleparameterne	58
5.3.1.	Accuracy.....	58
5.3.2.	Precision	59
5.3.3.	Recall.....	60
5.3.4.	F-Score	61
5.3.5.	Area under the curve	62
5.3.6.	Logarithmic Loss.....	63
5.4.	Sammenligning av metodene.....	64
5.5.	Oppsummering	65
6.	Konklusjon.....	67
6.1.	Kritikk til vårt forskningsdesign.....	68
6.2.	Videre forskning	69
7.	Litteraturliste.....	70

Oversikt over tabeller

Tabell 3.1 - Aksjeindekser inkludert som uavhengig variabel.....	19
Tabell 4.1 - Oversikt over inkluderte variabler	27
Tabell 4.2 - Deskriptiv statistikk fra datasettet	28
Tabell 5.1 - Vekting variabler BPM.....	39
Tabell 5.2 - Vekting ukedag BPM	40
Tabell 5.3 - Ytelse BPM.....	41
Tabell 5.4 - Vekting variabler Logistic Regression	42
Tabell 5.5 - Vekting ukedag Logistic Regression	43
Tabell 5.6 - Ytelse Logistic Regression	44
Tabell 5.7 - Vekting variabler SVM.....	45
Tabell 5.8 - Vekting ukedag SVM	46
Tabell 5.9 - Ytelse SVM	47
Tabell 5.10 - Sammenstilling vekting variabler	48
Tabell 5.11 - Sammenstilling vekting ukedag.....	49
Tabell 5.12 - Ytelse LD-SVM.....	51
Tabell 5.13 - Ytelse Boosted Decision Tree	52
Tabell 5.14 - Ytelse Decision Forest.....	54
Tabell 5.15 - Ytelse Decision Jungle	55
Tabell 5.16 - Ytelse Neural Network	57
Tabell 5.17 - Sammenstilling av modellenes totale ytelse	64

Oversikt over figurer

Figur 2.1 - Eksempel på beslutningstre	14
Figur 2.2 - Neural Network	16
Figur 3.1 - Eksport av varer fra Norge 2001 - 2017. Kilde: Statistisk Sentralbyrå (2018)	18
Figur 3.2 - Åpningstider på børser i UTC-tid.....	19
Figur 4.1 - Maskinlæring standard teknikk	22
Figur 4.2 - Prediksjonskategorier	31
Figur 4.3 - Eksempel på ROC-kurve.....	34
Figur 5.1 - Klassifiseringer BPM	40
Figur 5.2 - Klassifiseringer Logistic Regression.....	43
Figur 5.3 - Klassifiseringer SVM.....	46
Figur 5.4 - Klassifisering LD-SVM	50
Figur 5.5 - Klassifisering Boosted Decision Tree	52
Figur 5.6 - Klassifisering Decision Forest	53
Figur 5.7 - Klassifisering Decision Jungle	54
Figur 5.8 - Klassifisering Neural Network.....	56
Figur 5.9 - Sammenstilling Accuracy	58
Figur 5.10 - Sammenstilling Precision	59
Figur 5.11 - Sammenstilling Recall.....	60
Figur 5.12 - Sammenstilling F-score.....	61
Figur 5.13 - Sammenstilling AUC	62
Figur 5.14 - Sammenstilling Log loss	63

Oversikt over Appendiks

Appendiks A – Utdrag fra datasett uten normalisering

Appendiks B – Utdrag fra datasett med normalisering

Appendiks C – Oversikt over hyperparametere til modellene med forklaring

BEGREPSORDLISTE OG FORKORTELSER

AEX	Amsterdam Exchange Index
APT	Arbitrasjeprisingsteorien
AUC	Area Under the Curve.
BPM	Bayes Point Machine
CAC40	EuroNext Paris Benchmark indeks
DAX	Deutscher Aktienindex
EUR	Euro
FTSE100	Financial Times Stock Exchange 100 Index
GBP	Britiske pund
KVM	Kapitalverdimodellen
LD-SVM	Locally Deep Support Vector Machine
Log loss	Logarithmic Loss
LR	Logistic Regression
NASDAQ	NASDAQ-børsen i New York
NOK	Norske kroner
NYSE	New York Stock Exchange
OMX Copenhagen	Aksjeindeks bestående av de 20 mest omsatte aksjene på København-børsen
OMX Stockholm	Aksjeindeks bestående av de 30 mest omsatte aksjene på Stockholm-børsen
OSEBX	Oslo Børs Hovedindeks
ROC	Receiver Operating Characteristic
S&P500	Standard & Poor's 500 index
SSE Comp	Shanghai Stock Exchange Composite Index
SVM	Support Vector Machine
USD	Amerikanske dollar
VIX	Chicago Board Options Exchange Volatility Index

1. Innledning

Vi lever i en tid hvor den teknologiske utviklingen går svært raskt og vi har i dag maskiner som kan prosessere store mengder data på kort tid og til en marginal kostnad. I tillegg til at vi har tilgang på maskinkraft, så er også selve dataen lett tilgjengelig. Ved å inneha mye data så kan man analysere disse dataene for å oppdage mønster. Dette er mønster som, på grunn av mengden data, mennesker ikke evner å finne da vi har begrenset kapasitet. Det blir enda vanskeligere for mennesker å finne mønster når mengden data øker. Som mennesker tar vi ofte beslutninger basert på begrenset rasjonalitet. Det vil si at vi tar beslutningen når vi føler at vi har nok informasjon til å fatte den rette beslutningen. Nå som vi har muligheten til å prosessere større mengder data ved hjelp av maskinkraft så kan beslutningsgrunnlaget bli bedre, og følgelig er det naturlig å anta at beslutningene som fattes i større grad vil være riktige.

Maskinlæring er et underfelt innenfor kunstig intelligens, som åpner muligheter for å analysere store datasett for å finne mønster og lære fra disse mønstrene for deretter å generalisere med den hensikt å utvikle nye modeller. Modellene som blir utviklet skal kunne brukes for å vise hva som sannsynligvis vil skje i fremtiden. Det vil si at ved hjelp av maskinlæringsalgoritmer så kan man i noen tilfeller ta bedre beslutninger enn ved å kun bruke manuelle metoder.

Maskinlæring blir stadig en større del av samfunnet. Vi blir daglig eksponert for det, uten at vi nødvendigvis tenker igjennom at det er maskinlæring. Hver gang vi får en e-post så vil denne prosesseres av en algoritme som prøver å identifisere hvorvidt denne e-posten er spam eller ikke. Når du gjennomfører et søk på nettet så er resultatet du får basert på hvordan du har reagert på tidligere søk. I media er det også for tiden mye fokus på selvkjørende biler som er drevet av maskinlæring. Med andre ord så er maskinlæring rundt oss i utallige prosesser i det daglige og i et stadig voksende omfang. Vi ser også at det benyttes innenfor finansforskning, og har derfor lyst til å bruke maskinlæring som metode i vår studie. Vi ser derfor på denne masteroppgaven som en mulighet til å fordype oss i maskinlæring, samtidig som vi kan benytte oss av teori som vi har lært i løpet av MBA-studiet.

Norge er en liten økonomi som er avhengig av eksport og har historisk sett blitt påvirket av internasjonale forhold. Vi ønsker i vår masteroppgave å se på om det er mulig å bruke maskinlæring for å finne ut hvordan og i hvilken grad Oslo børs reagerer på endringer i internasjonale forhold.

I artikkelen «New Tricks for Econometrics» (2014) skriver Varian, H. R. om nye metoder innen økonometrien som gjør det å jobbe i Big Data-kontekster mer tilgjengelig. Det er teknikker som vi synes virker interessante og har derfor valgt å benytte noen av disse metodene i vår masteroppgave. Det finnes flere studier som har forsøkt å se på hvordan sammenhengen mellom Oslo børs og utvalgte makroøkonomiske variabler. Felles for mange av disse er at analysene er av tradisjonell økonometrisk karakter. Dette er modeller som for eksempel vektor autoregressiv analyse (VAR) (Gleditsch og Vollan (2012), Liavik og Andorsen (2013)), Principal Component Analysis (PCA) (Svarttjernet og Ulsrud (2016)) eller Ordinary Least Squares (OLS) (Brynstad og Johnsen (2016)). I vår studie har vi brukt metoder omtalt av Varian (2014) for å se om disse er egnet for å se på norske forhold.

1.1. Problemstilling

I denne studien har vi valgt problemstillingen:

I hvor stor grad kan vi ved hjelp av maskinlæring predikere om Oslo Børs, representert ut i fra hovedindeksen, vil gå opp eller ned i løpet av en dag ut i fra dagens endringer i utvalgte variabler?

Det finnes en rekke variabler som kan være med i en slik analyse. Siden Norge er en liten økonomi i internasjonal sammenheng med relativt store inntekter innenfor eksport har vi i stor grad valgt å fokusere på variabler som er relatert til handel med utlandet. Dette blir gjennomgått i kapittel 3 som omhandler valg av variabler.

Etttersom vi har valgt å bruke maskinlæring som metode for å generere prediktive modeller, er vi også avhengig av å ha mange observasjoner for å kunne både generere og teste modellen. Derfor har vi valgt å fokusere på variabler som har daglige observasjoner.

Vi vil igjennom denne studien bruke flere ulike maskinlæringsteknikker, og har dermed følgende underproblemstillinger:

Hvilke(n) maskinlæringsteknikker fungerer best for å predikere daglige bevegelser på Oslo Børs ut i fra våre valgte variabler?

Er det mulig, ved hjelp av maskinlæring, å si hvilke av våre utvalgte variabler som har størst sammenheng med Oslo Børs, representert ved Hovedindeksen?

1.2. Oppgavens oppbygging

Vi vil innlede studien med et todelt teorikapittel som tar for seg de teoretiske rammeverkene innenfor økonomi og maskinlæring som er nødvendig for å besvare problemstillingene våre. Vi vil deretter redegjøre for valg av variabler som vi som vi anser som potensielle faktorer som kan påvirke utviklingen på Oslo Børs hovedindeks.

Etter gjennomgangen av teori og variabler vil vi redegjøre for metodedelen av studien. Denne delen vil gjennomgå den trinnvise prosessen fra datainnsamling, datakonvertering, generering av maskinlæringsmodeller og hvordan resultatene kan analyseres. Vi vil også redegjøre for reliabiliteten og validiteten til studien vår.

Videre vil vi gå gjennom en analyse av de ulike maskinlæringsteknikkene med fokus på hvor god prediksjonsevne hver enkelt av disse har på vårt datasett. Etter analysen vil vi ha en diskusjon om hvordan dette kan knyttes opp til problemstillingene våre, etterfulgt av kritikk til metoden og forslag til videre forskning.

2. Teori

I dette kapitlet vil vi gå inn på teoretiske momenter som er viktige i vår besvarelse. Vi har valgt å dele teorikapitlet i to hoveddeler. Del en tar for seg økonomisk teori, og vil redegjøre for relevante teorier tilknyttet aksjemarkedet og prising av aksjer. Del to tar for seg maskinlæringsteori og vil redegjøre for maskinlæring overordnet og de teknikkene som vi har benyttet i våre analyser.

2.1. Økonomisk teori

I underkapitlene som omhandler ulike økonomiske teorier vil vi gi et innblikk i aksjemarkedet, og hva som driver prisene på markedet. Vi vil gå inn på noen modeller som forsøker å forklare prisingen av aksjer og vi vil redegjøre for andre teorier som har en annen innfallsvinkel.

2.1.1. Om aksjemarkedet

I aksjemarkedet omsettes aktiva, altså noe som har en verdi. Den største delen av denne handelen skjer på børser, men aksjer omsettes også utenfor børsene.

I Norges offentlige utredninger nr. 16, om Sparing og egenkapital 1995 lister Finansdepartementet (1995, s. 48) opp fem hovedoppgaver som aksjemarkedet har

Disse er, fritt fortalt,

- Bidra til at virksomheter får tilgang til egenkapital gjennom aksjekjøp. Denne egenkapitalen kommer i tillegg til opptjent egenkapital og kommer fra kapitaleiere som ønsker høyere avkastning enn de kan få ved passiv plassering
- Bidra til at kapital benyttes til verdiskapning. Økte investeringer i aksjemarkedet fører også til økt lånekapital
- Bidrar til at kapitaleiere kan spre risiko gjennom en diversifisert portefølje.
- Bidrar til spredning av kunnskap og kompetanse.
- Bidra til økt likviditet i virksomhetene gjennom muligheten for aksjeemisjoner.

2.1.2. Forventninger og usikkerhet

Økonomien har gjennom alle tider blitt forsøkt forklart gjennom omfattende studier med ulik innfallsvinkel. Felles for de alle er at det er umulig å lage en perfekt teori som tar høyde for alle mulige faktorer. Noen av faktorene som påvirker økonomien ligger i menneskers forventning og usikkerhet til den økonomiske utviklingen (Keynes, 1936). Det er variabler som er vanskelige å ta inn i en forklaringsmodell, men som kan skape forståelse for hvorfor markedene beveger

seg. Investorers vilje til å investere i risikable aktiva som f.eks. aksjer avhenger ifølge Keynes (1936) av hvilke forventninger investoren har til utviklingen. En investor som ikke har tro på utvikling i aksjemarkedet vil velge å ikke investere i aksjer, men heller plassere midlene sine i banken eller plassere de i obligasjoner med kort varighet. Dersom slike forventninger sprer seg blant investorer blir de dårlige forventningene selvoppfyllende, da flere investorer samtidig vil selge seg ut av sine aksjeposisjoner og kursene vil drives nedover. På den andre siden vil en felles forventning om gode tider i aksjemarkedet føre til økt etterspørsel og aksjeprisene vil dermed stige.

For aksjer så er det naturlig nok flere ting enn investorers framtidsutsikter som vil påvirke prisen. Vi vil i delkapitlene under redegjøre for teori som kan hjelpe oss å forstå hvordan markedet fungerer, og vi vil redegjøre for noen finansielle teorier som kan hjelpe investorer ta riktige beslutninger. Vi vil også redegjøre kort for random walk teorien som er en motvekt til disse.

2.1.3. Markedseffisiens

Et viktig tema å gå inn på når man skal vurdere hvorvidt en aksje er riktig priset er begrepet markedseffisiens. Det er et begrep som brukes for å beskrive i hvor sterk grad prisen er påvirket av informasjon. Markedseffisiens deles som oftest i tre typer; svak, halvsterk og sterk effisiens (Malkiel og Fama, 1970)

Med svak effisiens er den eneste informasjonen som reflekteres i prisen aksjens tidligere utvikling. Man vil ikke kunne oppnå ekstra gevinst ved å analysere tidligere bevegelser, da det ikke vil gi noen fordel i forhold til andre investorer ettersom informasjonen er tilgjengelig for alle.

Dersom man har en aksje med halvsterk effisiens så reflekterer prisen, i tillegg til aksjens utvikling, markedets utvikling for øvrig, politiske beslutninger, nyheter og all tenkelig offentlig tilgjengelig informasjon.

Ved sterk effisiens så reflekterer prisen all informasjon, også innsideinformasjon. Det vil si at aksjen er riktig priset og at den eneste potensielle gevinsten ligger i betaling for risikoen man er villig til å ta. Fullstendig effisiens vil ifølge Bodie et al. (2007, s. 361) aldri kunne oppnås, da det alltid vil være noen på innsiden av en bedrift som har tilgang til mer informasjon enn markedet for øvrig og således kan se en mulig prisutvikling før markedet.

2.1.4. Kapitalverdimodellen KVM

Kapitalverdimodellen (KVM) er en sentral modell innen finansteorien, se Sharpe (1964) og Lintner (1965). KVM benyttes for å beregne forventet avkastning til en aksje eller aksjeportefølje j hvor risikopremien β er en sentral komponent.

KVM uttrykkes slik:

$$E(r_j) = r_f \times (1 - s) + \beta_j \times [E(r_m) - r_f \times (1 - s)] \quad (2.1)$$

Hvor:

- $E(r_j)$ er forventet avkastning
- Risikofri rente er r_f med fratrekk for skatt $(1 - s)$
- Aksjens (j) risikopremie er β_j .
- Forventet markedsrente er $E(r_m)$.

Risikopremien β_j kalles også for relativ markedsrisiko, og kan uttrykkes på følgende måte.

$$\beta_j = \frac{Kov(r_j, r_m)}{Var(r_m)} \quad (2.2)$$

Kovariansen (Kov) mellom forventet avkastning for aksje j og forventet markedsrente delt på markedets varians (Var) gir altså forventet risikopremie. Fra formel (2.2) kan vi lese at en risikopremie større enn 1 vil gi en bevegelse i samme retning som markedet, men med større effekt. Den er med andre ord mer volatil enn markedet som helhet. Hvis det var tilfellet så vil en investor kreve høyere avkastning for en slik aksje ettersom han tar større risiko enn ved å investere i markedet som helhet. Dersom risikopremien er mellom 0 og 1 så vil risikoen ved å investere i aksjen være mindre enn markedet som helhet, da svingningene er mindre, og således gir lavere gevinst eller tap.

2.1.5. Arbitrasjeringsteorien

Arbitrasjeringsteorien (APT) er en alternativ modell til KVM, se Ross (1976). APT benytter summen av flere faktorer for å finne avkastning til en aksje eller aktivum ved å se på forventet avkastning og aktivumets følsomhet for endring i en eller flere faktorer. Det kan være for eksempel makroøkonomiske faktorer som f.eks. inflasjon eller brutto nasjonalprodukt. Det kan også være faktorer tilknyttet aksjemarkedet, f.eks. en aksjeindeks. Her kan man ekskludere eller

inkludere flere faktorer ut i fra hva man vet eller tror påvirker avkastningen til aksjen eller aktivum. Formelen for en multifaktormodell ser slik ut:

$$r_i = E(r)_i + \beta_{i_1}F_1 + \beta_{i_2}F_2 + \dots + \beta_{i_n}F_n + e_i \quad (2.3)$$

Hvor r_i er avkastning på aksjen i , $E(r)_i$ er forventet avkastning til aksje i , $\beta_{i_1}F_1$ er et uttrykk for aksjen i sin sensitivitet for faktor F_1 , og e_i er idiosynkratisk risiko, det vil si særegen risiko for aksjen i .

2.1.6. Random walk

Som motvekt til KVM og Arbitrasjepricingsteorien så finner vi Random Walk-teorien som ble publisert av Burton Malkiel på 70-tallet. Den har sitt opphav i artikkelen «Some A Posteriori Probabilities in Stock Market Action» publisert av Alfred Cowles 3rd og Herbert E. Jones (1937). De fant i sin forskning ut at aksjers avkastning er tilfeldige, og tidligere aksjekurs ikke har noe å si for kommende aksjekurs.

Dersom denne teorien stemmer så vil det ikke være mulig å predikere bevegelser på Oslo Børs, men den teknologiske utviklingen har gått videre siden forskningen ble publisert, og vi har nå muligheten til å undersøke aksjeprisbevegelser i en mye større kontekst enn tidligere.

2.2. Maskinl ring

I dette delkapittelet vil vi f rst p  inn p  hva maskinl ring er. Deretter vil vi g  inn p  noen teoretiske momenter rundt hvordan man m ler hvor *god* en maskinl ringsmetode fungerer mot et datasett. Til slutt vil vi g  gjennom grunnleggende teoretiske momenter til hver av de ulike teknikkene som vi har valgt   benytte.

2.2.1. Definisjon

Maskinl ring er et sett av verkt y som kan brukes for   gi st rre innsikt i data. Maskinl ringsprosessen handler om   bygge statistiske modeller som kan predikere eller estimere en avhengig variabel ut i fra et sett av uavhengige variabler (James et. al. 2017, s. 15). Disse verkt yene kan bli klassifisert som st ttende eller ikke st ttende l ring. Ved st ttende l ring blir den avhengige variabelen markert med korrekt resultat og gir f ringen for hvordan l ringen skal skje. Med ikke-st ttende l ring blir ikke den avhengige variabelen merket, og l ringsprosessen g r ofte ut p    gruppere data.

Maskinl ring kan enkelt forklares som at vi har et svar Y og p antall ulike uavhengige variabler, ofte kalt prediktorer, X_1, X_2, \dots, X_p . Dette kan skrives p  den generelle formen:

$$Y = f(X) + \epsilon \quad (2.4)$$

Her er f en fast, men ukjent funksjon av X_1, \dots, X_p , og ϵ er et tilfeldig feilledd som er uavhengig av X og har et gjennomsnitt p  0. I denne formelen representerer f den systematiske informasjonen som X gir oss p  Y .

Som nevnt innledningsvis  nsker vi   bruke maskinl ring til predikering. I mange situasjoner er verdiene til de uavhengige variablene X lett tilgjengelige mens svaret Y ikke er det. I denne situasjonen, siden feilleddet har et gjennomsnitt p  0, kan vi predikere Y ved bruk av

$$\hat{Y} = \hat{f}(X) \quad (2.5)$$

Hvor \hat{f} representerer v rt estimat for f og \hat{Y} representerer resultatet av predikeringen av Y . Ofte vil funksjonen av \hat{f} v re vanskelig   visualisere og/eller forklare og blir dermed ofte referert som en svart boks. Dette gj r at man n dvendigvis ikke bryr seg om den eksakte formen av \hat{f} gitt at den gir gode predikeringer for Y .

Variablene kan bli klassifisert som enten *kvantitative* eller *kvalitative* (også kalt *kategoriske*) (James et. al. 2017, s. 28-29). Kvantitative verdier er verdier som er numeriske som for eksempel en valutakurs. Kvalitative verdier er verdier som kan settes inn i K klasser (også kalt kategorier). Eksempler på kategoriske verdier kan være ukedag eller kjønn. Hvilken type variabel Y er avgjør hvilken type maskinlæringsproblem som skal løses. Hvis Y er en kvantitativ verdi vil problemet som skal løses klassifiseres som et regresjonsproblem, mens i de tilfellene hvor Y er en kvalitativ variabel kalles det et klassifiseringsproblem. Innenfor samme datasett kan de uavhengige variablene X_1, \dots, X_p være både kvantitative og kvalitative, uavhengig av om det er et kvalitativt eller kvantitativt problem som skal løses. Det vil være ulike typer maskinlæringsalgoritmer som kan løse ulike typer problemer. Hver maskinlæringsalgoritme har ulik grad av fleksibilitet. Lineære modeller er generelt sett mindre fleksible enn ikke-lineære modeller. I vår studie ønsker vi å se på både lineære og ikke-lineære kvalitative metoder innenfor maskinlæring ved hjelp av støttende læring.

2.2.2. Måling av ytelse

Innen maskinlæring er det mange ulike tilnærminger for å analysere et problem. Det vil ofte være nødvendig å bruke flere forskjellige teknikker for å analysere samme problem. Grunnen til dette er at det er «no free lunch in statistics» (James et. al. 2017, s. 29); det er ingen enkelt metode som bedre enn alle de andre metodene over alle mulige datasett. Det vil derfor ofte være nødvendig å bruke flere ulike teknikker og sammenligne dem med hverandre.

Den enkleste metoden for å kvantifisere treffsikkerheten på \hat{f} er å se på treningsfeilraten til modellen. Ved bruk av en klassifiseringsmodell måler vi treningsfeilraten til modellen (James et. al. 2017, s. 37). Treningsfeilraten er andelen feil predikering som blir gjort når vi bruker vårt estimat på \hat{f} på treningsobservasjonene:

$$\frac{1}{n} \sum_{i=1}^n I(y_i \neq \hat{y}_i) \quad (2.6)$$

Her er \hat{y}_i en predikert klassifisering for den i -ende observasjonen ved bruken av \hat{f} . $I(y_i \neq \hat{y}_i)$ er en binær verdi som er 1 hvis $y_i \neq \hat{y}_i$ og 0 hvis $y_i = \hat{y}_i$. Hvis $I(y_i \neq \hat{y}_i) = 0$ på en observasjon var prediksjonen på klassifiseringsmodellen korrekt.

Formel (2.6) refererer til hvor godt klassifiseringsmodellen gjør det mot treningsdata. Det vi er mer interessert i er hvordan modellen gjør det mot data som ikke ble brukt for å beregne \hat{f} (testdata). Testfeilraten på et datasett med testobservasjoner på formen (x_0, y_0) er gitt med

$$Ave(I(y_0 \neq \hat{y}_0)) \quad (2.7)$$

Hvor \hat{y}_0 er den predikerte klassifiseringen som resulterer i å teste modellen på en testobservasjon med predikator x_0 . En *god* klassifiseringsmodell er en som har en testfeilrate som er så liten som mulig. Hvis vi har en modell som har lav feilrate mot testdata refererer vi ofte til en modell som generaliserer godt. Dette vil være en modell som kan med *god* ytelse predikere observasjoner som enda ikke har inntruffet (ny data).

Når man drøfter om hvor godt en maskinlæringsmodell lærer og generaliserer er det to uttrykk som ofte diskuteres, dette er overtilpassing og undertilpassing. Overtilpassing og undertilpassing er ofte årsaken til dårlig ytelse på modeller opprettet ved hjelp av maskinlæring.

Overtilpassing refererer til når en maskinlæringsalgoritme modellerer treningsdata for godt (James et. al. 2017, s. 22). Dette skjer når en modell lærer detaljer og behandler støy som relevant informasjon i slik utstrekning at det negativt påvirker modellen sin ytelse på nye data. Problemet med overtilpassing er at tilfeldige svingninger og støy legges inn som konsepter i modellen. Dette kan være observasjoner som ikke gjelder ny data og påvirker modellens evne til å generalisere negativt. Overtilpassing er et problem som oftere oppstår på ikke-lineære modeller da de har større fleksibilitet enn lineære modeller.

Undertilpassing refererer til når en modell behandler faktisk informasjon som støy, og dermed ikke evner å generalisere data på en god måte. En undertilpasset modell vil ofte ha dårlig ytelse når den testes på nye data. Dette vil gi utslag i at maskinlæringsmodellen vil få dårlige resultater både mot trenings- og testdata.

Ideelt sett vil man ha en modell som er en plass midt i mellom en overtilpasset og undertilpasset modell. Dette målet er ofte vanskelig å oppnå i praksis. For å unngå at maskinlæringsmodellene som blir generert blir overtilpasset eller undertilpasset finnes det ulike teknikker avhengig av hvilken maskinlæringsteknikk som benyttes. Ofte vil dette handle om å begrense modellene med ulike parametere for å unngå at modellene for mange detaljer.

2.2.3. Varians vs skjevhet

Varians refererer til hvor mye \hat{f} forandrer seg hvis vi trener opp en modell ved bruk av et annet datasett enn det opprinnelige (James et. al. 2017, s. 34-35). Siden treningsdata blir brukt for å tilpasse en modell, vil en annen treningsdata vil gi en annen \hat{f} . Ideelt sett ønsker vi at estimatet vårt for f ikke skal variere mye mellom ulike treningsdatasett. Hvis en metode har høy varians vil små endringer i treningsdata resultere i store endringer i \hat{f} . Generelt sett har fleksible metoder høyere varians enn metoder med lav fleksibilitet.

Skjevhet referer til et feilledd som blir introdusert i modellen ved å kalkulere et virkelig problem, som kan være kompleks, med en forenklet modell (James et. al. 2017, s. 35). Eksempelvis hvis vi bruker en lineær funksjon for å finne ut forholdet mellom Y og X_1, X_2, \dots, X_p , og dette forholdet ikke er lineært, vil dette resultere i høy skjevhet. Generelt sett har fleksible metoder lavere skjevhet enn metoder som er mindre fleksible.

2.2.4. Utelatt variabel-skjevhet

Utelatt variabel-skjevhet er et kjent problem i økonomisk forskning, se Clarke (2005). Problemet kan oppstå når en modell har utelatt en eller flere viktige variabler. Når enkelte viktige variabler blir utelatt, vil modeller kunne lage skjevhet ved å overestimere eller underestimere inkluderte variabler. Problemet rundt utelatte variabler er spesielt stort i analysen av økonomiske markeder da det er utrolig mange variabler som kan påvirke markedet.

2.2.5. Maskinlæringsalgoritmer

Avhengig av type problem som skal løses er det ulike maskinlæringsalgoritmer som er mer egnet enn andre. Vi har valgt å dele maskinlæringsalgoritmene i to kategorier, lineære og ikke-lineære. De lineære algoritmene vi har valgt å benytte er Bayes Point Machine, Logistic Regression og Support Vector Machine. De ikke-lineære algoritmene vi har valgt å benytte er Locally-Deep Support Vector Machine, Boosted Decision Tree, Decision Forest, Decision Jungle og Neural Network. I dette delkapittelet vil vi gå kort igjennom de teoretiske momentene tilknyttet hver algoritme.

2.2.5.1. *Bayes Point Machine*

Bayes Point Machine (BPM) er en algoritme som bruker en bayesiansk tilnærming til lineær klassifisering, se Qi et.al (2001). Algoritmen tilnærmer seg tilnærmet det teoretisk optimale bayesianske gjennomsnittet av en gruppe lineære klassifiserere (i form av generaliseringsytelse) ved å velge en «gjennomsnittlig» klassifiserer, Bayes Point.

Siden BPM er en bayesiansk klassifikasjonsmodell er den mindre utsatt for overtilpassing mot treningsdata.

2.2.5.2. *Logistic Regression*

I Logistic Regression (LR) blir de uavhengige input variablene kombinert lineært ved bruk av vekt og koeffisienter (referert som beta verdier) for å kalkulere en avhengig variabel Y (James et. al. 2017, s. 130-132). Hovedforskjellen mellom LR og lineær regresjonsanalyse er at Y som skal kalkuleres er kvalitativ, mens lineær regresjonsanalyse prøver å finne en numerisk (kvantitativ) verdi.

Målet med LR er å finne den beste fornuftige tilpasningen til en modell ved å beskrive forholdene mellom de uavhengige variablene og den avhengige variabelen. Maskinlæring bruker LR for å generere koeffisientene (og dens standardavvik og signifikansnivåer) på en formel for å forutsi en logit-transformasjon med sannsynlighet for tilstedeværelse av karakteristikk av interesse.

LR antar at det er en logisk distribusjon av observasjonene, hvor sannsynligheten for at et utfall tilhører klasse 1 er formelen:

$$p(X) = \beta_0 + \beta_1 X_1, \dots, \beta_n X_n \quad (2.8)$$

Hvor:

- X er en n -dimensjonal vektor som inneholder verdiene av alle featurene i forekomsten.
- p er en logistisk distribusjonsfunksjon.
- $\beta_0 + \beta_1 X_1, \dots, \beta_n X_n$ er de ukjente parameterne for logistikkfordelingen.

Algoritmen forsøker å finne de optimale verdiene for $\beta_0 + \beta_1 X_1, \dots, \beta_n X_n$ ved å maksimere log-sannsynligheten for parameterne gitt inngangene.

2.2.5.3. *Support Vector Machine*

Support Vector Machines (SVM) er en maskinlæringsalgoritme som vi kan bruke til klassifisering (James et. al. 2017, s. 337-338). Ved hjelp av merket data plotter algoritmen data inn i et p -dimensjonalt rom (hvor p er antallet uavhengige variabler i datasettet) for å så kalkulere et hyperplan i dette området. Et hyperplan er definert som en vektor hvor

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p = 0 \quad (2.9)$$

For parameterne $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ og de uavhengige variablene $X = (X_1, X_2, \dots, X_p)^T$. Vektoren vil dele det p -dimensjonale rommet i to ved binær klassifisering, hvor en klasse befinner seg på hver side av denne vektoren. Hvis en ny X ikke oppfyller (2.9) men i stedet oppfyller

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p > 0 \quad (2.10)$$

befinner X seg på den ene siden av hyperplanet og kan tilordnes den ene av de 2 klassene. På den andre siden hvis

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p < 0 \quad (2.11)$$

befinner X seg på den andre siden av hyperplanet, og klassifiseres som den i den andre klassen.

2.2.5.4. *Locally-Deep Support Vector Machine*

Locally-Deep Support Vector Machine (LD-SVM) er utviklet som et forsøk på å fremskynde en ikke-lineær SVM prediksjonsmodell (Jose et. al. 2013). LD-SVM lærer en trebasert funksjon som er høydimensjonal og sparsom. Ved bruk av trestrukturerte funksjoner introduserer LD-SVM ikke-lineariteter i modellene i motsetning til SVM. I følge Jose et. al (2013) vil denne tilnærmingen være raskere i både treningstid og prediksjonstid, samtidig som prediksjonsytelsen er *god*.

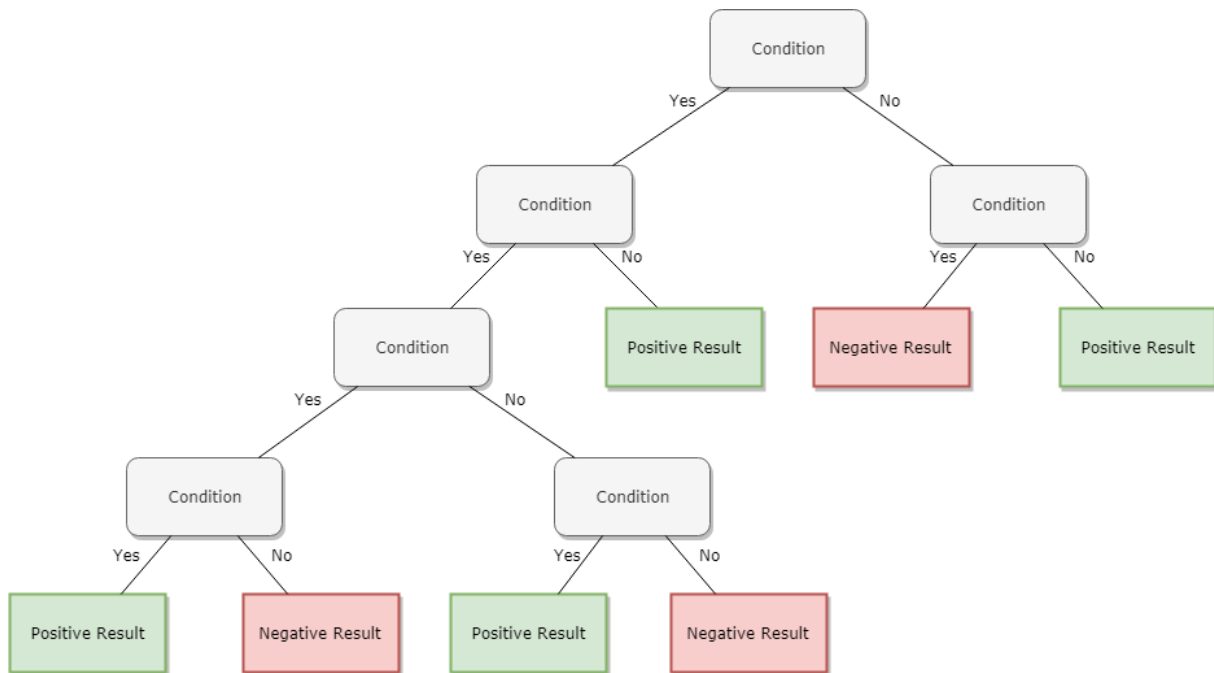
En LD-SVM lærer prediksjon på formen:

$$y(x) = \text{sign}\left(\sum_k p(w_k|x)w_k^t \phi_k(x) + b\right) \quad (2.12)$$

Hvor $p(w_k|x)$ er en ikke-negativ funksjon som kan bli sett på sannsynligheten for å velge en klassifikator w_k for et gitt punkt x . En kombinasjon av klassifikatorer w_k og variabler ϕ_k er valgt for hver x .

2.2.5.5. *Boosted Decision Tree*

Boosted Decision Tree er en maskinlæringsalgoritme som konstruerer et beslutningstre ved hjelp av støttet læring (James et. al. 2017, s. 311). Et beslutningstre er en rettet asyklisk graf hvor hver node i beslutningstreet inneholder en test med ulike utfall. De ulike utfallene av denne testen danner en gren som går til en ny node eller en sluttnode. En sluttnode er en node som har et utfall, som i vårt tilfelle vil være en klassifisering 1 eller 0, og beregningen i beslutningstreet slutter når denne noden nås.



Figur 2.1 - Eksempel på beslutningstre

Ved å bruke en teknikk som kalles boosting genererer maskinlæringsalgoritmen flere beslutningstrær sekvensielt, hvor det første treet blir brukt som utgangspunkt i det neste (James et. al. 2017, s. 321-323). I hver iterasjon vil beslutningstreet bli testet mot tilfeldig tildelte utgangspunkter, med et såkalt ulikt random seed. Et random seed er et ikke-negativt heltall som brukes for å initiere en tilfeldig sekvens av observasjoner som skal brukes. Ved å bruke et ulikt random seed ved testing av beslutningstrær vil fordelingen av data være litt forskjellig for hver iterasjon. Hver iterasjon vil gjøre tilpasninger på beslutningstreet for å bedre modellen. Etter en rekke iterasjoner vil man til slutt stå igjen med en modell som inneholder et enkelt beslutningstre.

2.2.5.6. *Decision Forest*

Maskinlæringsteknikken Decision Forest bruker også beslutningstrær for å generere en modell, men har ikke som mål å generere kun et beslutningstre slik som Boosted Decision Tree. Treningsalgoritmen for Decision Forest bruker en teknikk kalt bagging (James et. al. 2017, s. 316-317). Ut fra et gitt treningssett $X = X_1, \dots, X_n$ med klassifisering $Y = Y_1, \dots, Y_n$, så vil bagging velge en tilfeldig prøve av treningssettet B antall ganger og generere et beslutningstre fra hver av disse prøvene:

For $b = 1, \dots, B$:

1. Ta en tilfeldig prøve av treningsdatasettet $X, Y; X_b, Y_b$
2. Tren et binært beslutningstre f_b på X_b, Y_b .

Etter trening sitter man igjen med flere ulike beslutningstrær. Når predikeringer skal gjøres på ny data x' tar vi gjennomsnittet av predikeringene fra hvert individuelle beslutningstre på x' :

$$\hat{f} = \frac{1}{B} \sum_{b=1}^B \hat{f}_b(x') \quad (2.13)$$

Bagging kan lede til bedre modellytelse siden det reduserer variansen til modellen, uten å øke skjevhet, noe som reduserer overtilpassing. Predikeringene til et enkelt beslutningstre vil være veldig sensitivt til støy i treningsdatasettet, mens gjennomsnittet av flere beslutningstre nødvendigvis ikke er dette så lenge beslutningstreene ikke er korrelert med hverandre. Ved å bruke bagging og trene beslutningstre på ulike deler av datasettet vil vi redusere korrelasjon mellom beslutningstreene.

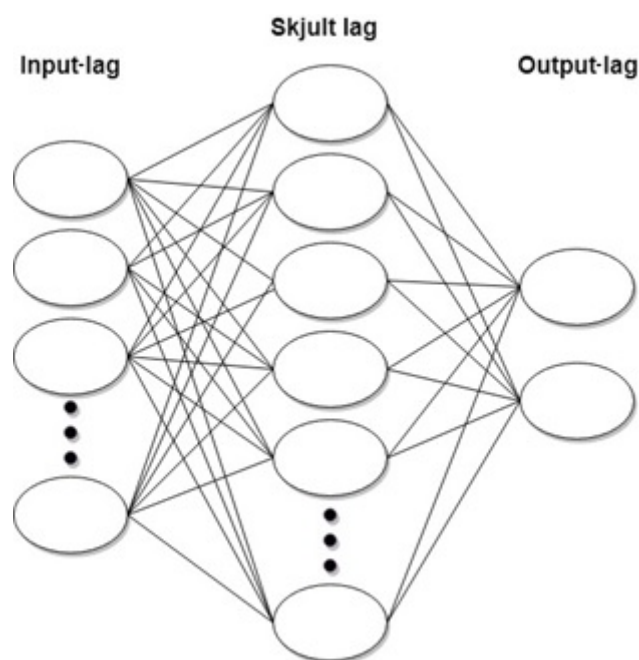
2.2.5.7. *Decision Jungle*

En Decision Jungle er en utvidelse av Decision Forest som er beskrevet i forrige kapitlet, se Shotton et. al. (2013). Den bygger opp et ensemble av beslutningstrær, men følger ikke de samme reglene. I motsetning til konvensjonelle beslutningstrær som bare tillater en bane til hver node, tillater en Decision Jungle flere baner fra roten til hvert blad. I følge Shotton et. al. vil som oftest en Decision Jungle kreve mindre minne og i enkelte tilfeller også generalisere bedre enn Decision Forest som følger de konvensjonelle reglene. Ulempen med denne algoritmen er at treningstiden er betydelig lenger.

2.2.5.8. *Neural Network*

Et Neural Network er inspirert av hvordan den menneskelige hjernen fungerer (Mund 2015, s. 7). Algoritmene bygger et nettverk av beregningsenheten som kalles noder eller nevroner. I et typisk nettverk er det tre lag av noder – input-laget, det skjulte laget og output-laget. Neural Network som har flere enn tre lag kalles for nettverk som bruker dyp læring. Slike nettverk har mer enn ett skjult lag.

Input-laget består av en node per uavhengig variabel som skal brukes i analysen og holder på denne verdien. Hver eneste node kobles til hver node i neste lag og sender denne verdien inn som input. Det skjulte laget inneholder en rekke noder som tar inn denne vektete samlingen av input fra det foregående laget og produserer en output gjennom en aktiveringsfunksjon. Disse nodene, med ulike aktiveringsfunksjoner, blir konstruert gjennom en læringsalgoritme fra datasettet. Det siste laget i et Neural Network er output-nodene. Siden vi konstruerer et binært Neural Network, vil dette laget ha to noder. Output-nodene tar imot vektete inputs fra alle nodene i det foregående laget og beregner den veide summen av alle disse verdiene. Gjennom denne funksjonen kan nettverket gjøre en sannsynlighetsberegning for utfall på hver node ut i fra ny input som kommer inn i input-laget.



Figur 2.2 - Neural Network

3. Variabler

I dette kapitlet vil vi redegjøre for hvilke variabler vi vil inkludere i vårt forskningsdesign, samt begrunne utvalget.

Uavhengig av forskningsdesign så vil man alltid ha behov for data for å kunne gjennomføre forskning. Det er vanlig å skille mellom primærdata og sekundærdata, hvor primærdata er data som forskeren selv har produsert gjennom f.eks. intervjuer, mens sekundærdata er data som allerede finnes og som kan hentes ut fra for eksempel databaser.

I vår masteroppgave har vi benyttet data som allerede finnes, altså sekundærdata. Det som da er viktig for oss er at vi har hentet data fra troverdige kilder, samt at vi har behandlet den på en måte som ikke endrer dens karakteristikker. Vi har hentet ut data fra kildene Oslo Børs, Thomson Reuters, Norges Bank, Statistisk sentralbyrå, Yahoo Finance og U.S. Energy Information Administration. Alle disse anser vi som troverdige kilder både fordi de er anerkjente leverandører, og fordi de figurerer som kilder i flere andre forskningsprosjekt.

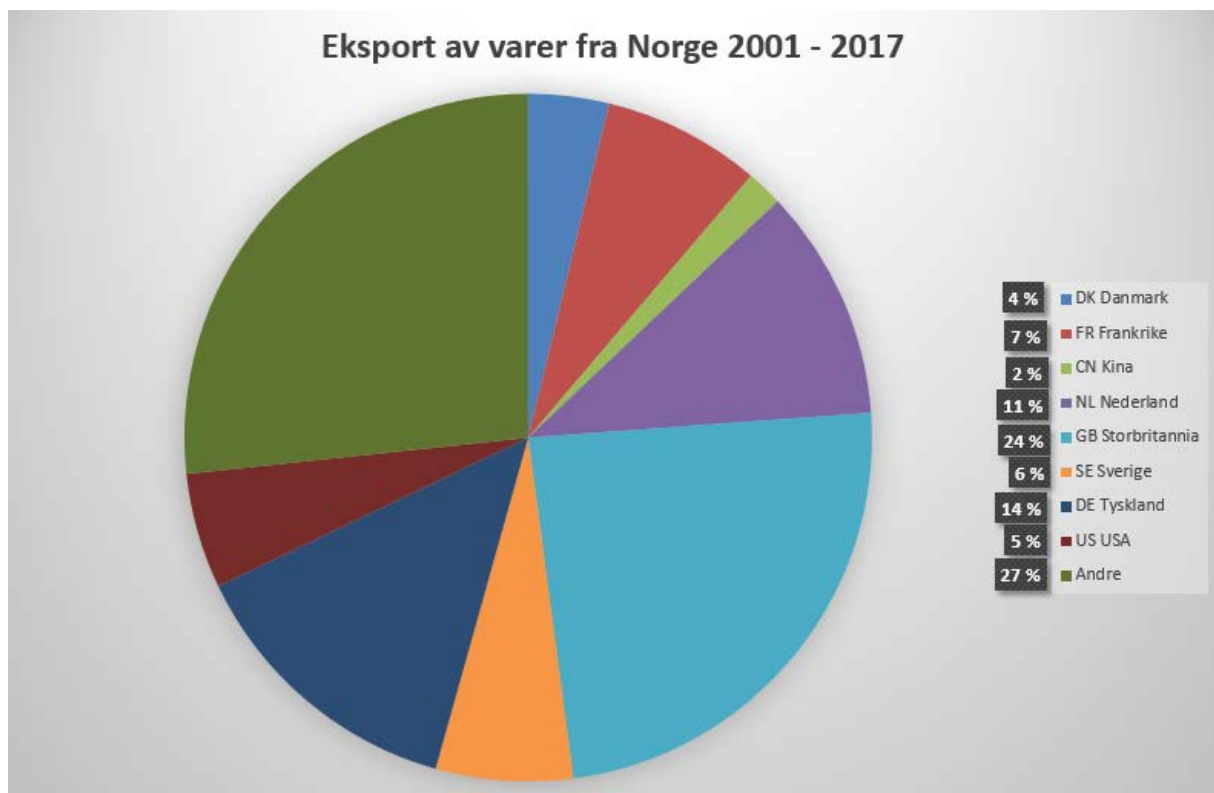
Vi benytter oss av daglige observasjoner i perioden 2001-2017. Ettersom det er varierende helligdager i de nevnte land, og det ikke finnes data fra lørdag og søndag så må vi behandle dataen slik at vi får verdier på samme dato for alle variabler. Det er også ulike tidspunkt for notering av variablenes verdi. Dette er knyttet til både børsenes åpningstid og tidssoner. Vi vil komme tilbake til hvordan vi har håndtert dette i kapittel 4.1.

3.1. Avhengig variabel

Vi ønsker å gi svar på om Oslo Børs skal opp eller ned basert på et gitt sett variabler som vi antar har forklaringskraft. Vi har valgt å bruke Oslo Børs hovedindeks (OSEBX) som den avhengige variabelen i vårt prosjekt, da vi anser den som representativ for børsen som helhet. Den består av det som Oslo børs (www.oslobors.no, 2018) omtaler som «et representativt utvalg av alle noterte aksjer på Oslo Børs». Den består for tiden av 67 aksjer. Indeksen skal være investerbar, med andre ord så skal det være mulig å både kjøpe og selge andeler med relativt lave transaksjonskostnader. Ettersom Oslo Børs er et regulert marked gjennom bl.a. Lov om verdipapirhandel (2007) så ansees den som et trygt marked for å handle verdipapirer.

3.2. Uavhengige variabler

Som nevnt innledningsvis er den Norske økonomien liten sammenlignet med verdensøkonomien, og den er åpen. Det er derfor naturlig å anta at Norge påvirkes av verden, og ikke motsatt. Vi har derfor valgt å se på hvem som er Norges største handelspartnere med hensyn på eksport av varer når vi skal velge uavhengige variabler. Vi har valgt å se bort fra eksport av tjenester, da omfanget av tjenesteeksport er forsvinnende liten sammenlignet med vareeksport. Tall fra SSB viser at Norge i perioden 2001 - 2017 eksporterte varer for i nærmere 13 000 mrd. kroner. Våre 8 største handelspartnere står for 73,31 % av denne omsetningen. Det er i hovedsak europeiske land, men også USA og Kina – to land som ofte brukes for å forklare og analysere trender i verdensøkonomien.



Figur 3.1 - Eksport av varer fra Norge 2001 - 2017. Kilde: Statistisk Sentralbyrå (2018)

Vi har valgt å benytte et bredt utvalg uavhengige variabler for å forsøke å predikere utviklingen på Oslo Børs. Noen av variablene har vært benyttet i lignende studier tidligere. For eksempel mener Svartjernet og Ulsrud (2016) at Storbritannias aksjemarked påvirker Oslo Børs. I samme studie presenteres også at oljepris og Kinas børs har signifikans for Oslo Børs sin utvikling.

Utvalget vårt inkluderer åtte aksjeindekser fra våre største handelspartnere. Data for indeksene er hentet fra Thomson Reuters databasen. Felles for indeksene vi har valgt er at vi mener de

representerer aksjemarkedet i det respektive landet godt og derfor er gode indikatorer på den økonomiske temperaturen.

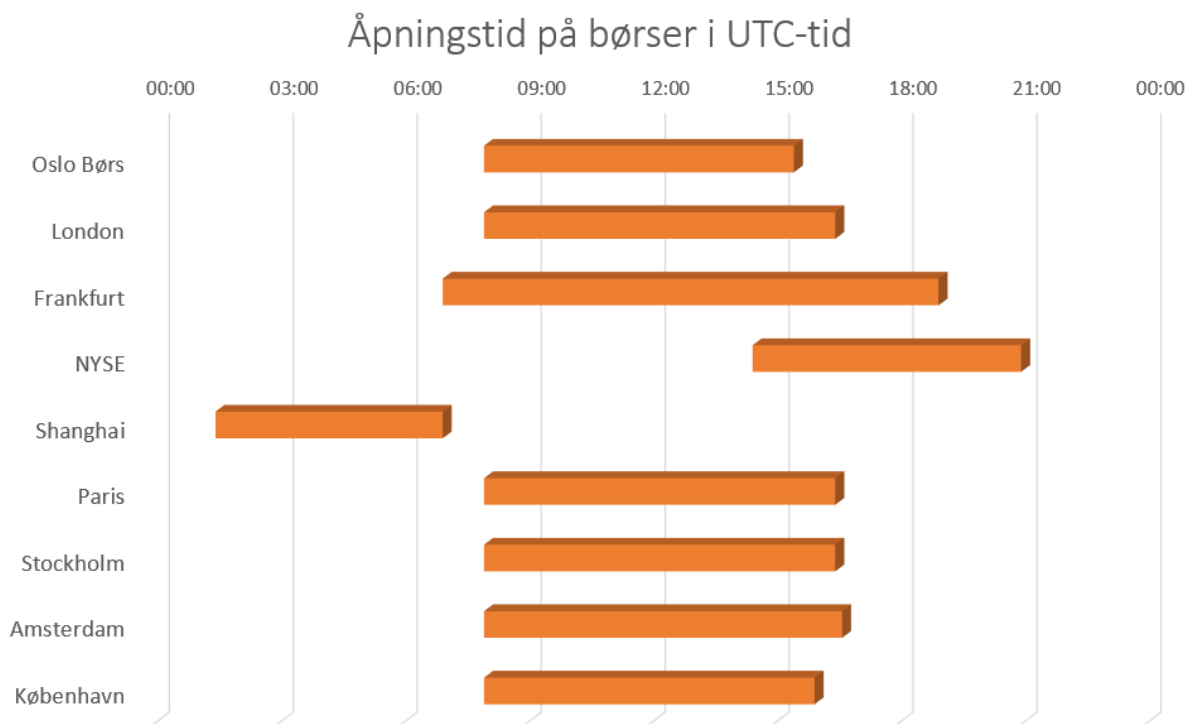
Vi har også valgt å inkludere OSEBX noteringen fra dagen før, da det kan være interessant å se om det er noen sammenheng mellom bevegelse fra den ene dagen til neste. Data for OSEBX er hentet fra Oslo Børs.

Følgende aksjeindekser har blitt inkludert:

Indeks	Børs	Land	Valuta
FTSE100	London Stock Exchange	Storbritannia	GBP
DAX	Frankfurter Wertpapierbörse	Tyskland	EUR
SPCOMP	NYSE/NASDAQ	USA	USD
SSE	Shanghai Stock Exchange	Kina	CNY
CAC40	Euronext Paris	Frankrike	EUR
SWE OMX	Stockholm Stock Exchange	Sverige	SEK
DK OMX	Copenhagen Stock Exchange	Danmark	DKK
AEX	Euronext Amsterdam	Nederland	EUR
OSEBX	Oslo Børs	Norge	NOK

Tabell 3.1 - Aksjeindekser inkludert som uavhengig variabel

De forskjellige børsene har ulik åpningstid. Som man ser av Figur 3.2 så har de fleste børsene nokså lik åpningstid som Oslo Børs, men Shanghai og NYSE/NASDAQ har naturlig nok åpningstid som i stor grad faller utenfor den norske.



Figur 3.2 - Åpningstider på børser i UTC-tid

Som et mål på volatilitet i markedet har vi valgt å inkludere volatilitetsindeksen VIX. Den er et mål på forventede svingninger for S&P500 de neste 30 dager, omregnet til årlig proSENTSATS. Hvis VIX indeksen har en høy prosentverdi kan dette tyde på stor usikkerhet i markedet. I perioder med stor usikkerhet vil muligens investorer handle annerledes enn hvis det lav usikkerhet. Dette vil potensielt påvirke svingningene på de ulike børsene. VIX-verdiene er hentet fra Yahoo Finance.

Den neste variabelen vi har inkludert er oljepris. Norge er i stor grad avhengig av oljeeksport og således er oljepris en faktor som påvirker børsen (Henriksen og Killingstad, 2013). Flere av selskapene som inngår i OSEBX er direkte eller indirekte involvert i handel med olje. Den norske oljen omsettes som Brent-olje på Intercontinental Exchange (ICE). Den daglige spotprisen tilsvarer sluttnoteringen til Brent Crude på Londonbørsen, som er klokken 17:30 britisk tid (UTC +0 vintertid / UTC +1 sommertid). Oljeprisene er hentet fra U.S. Energy Information Administration.

Vi har også valgt å inkludere de valutakursene som hovedsakelig benyttes til betaling av Norges eksportvarer. Valutakursen er prisen på utenlandsk valuta uttrykt i norske kroner, og styres av tilbud og etterspørsel i valutamarkedet. Tilbud og etterspørselen i valutamarkedet er styrt av transaksjoner med utlandet i forbindelse med handel av varer og tjenester, kapitaltransaksjoner og stønader. Valutakursene vi har valgt å inkludere er USD, EUR og GBP. Kursene er hentet fra Norges Bank.

USD er valgt fordi det er valutaen som benyttes til oljekontrakter. Det er derfor naturlig å anta at norske selskaper med driftsutgifter i NOK og inntekter i USD vil kunne påvirkes av forholdet mellom NOK og USD. EUR er valutaen som benyttes til størstedelen av vår eksport. Omtrent 45 % av varer eksportert i perioden 2001-2017 gikk til land som benytter EUR som valuta. GBP er tatt med siden det er valutaen som benyttes hos vår største handelspartner, Storbritannia. Riktignok er mye av handelen med Storbritannia knyttet til oljekontrakter, som omsettes i USD, men GBP er også en stor valuta i verdenssammenheng og har hatt store variasjoner mot USD.

Vi har også valgt å ta med ukedag som variabel da tidligere studier ved f.eks. Holm (2007) sier at det er sammenheng mellom ukedag og børsutvikling. Studiet viser at gjennomsnittlig avkastning til hovedindeksen er ulik avhengig av hvilken dag i uken det er. Holm har blant annet funnet signifikant positiv meravkastning på fredager enn den gjennomsnittlige handelsavkastningen.

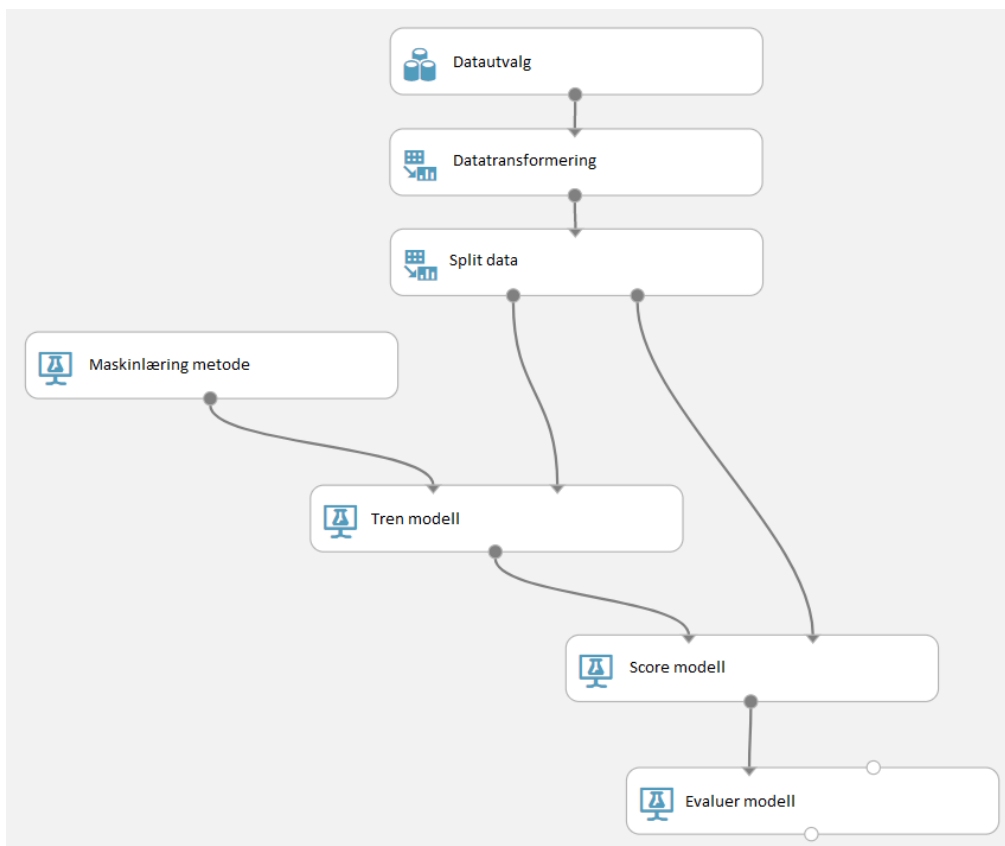
Likviditet i markedet er også en variabel som kan tenkes å påvirke børsen. Som et mål på likviditet har vi brukt daglig omsetning på Oslo Børs målt i NOK. Verdiene er hentet fra Oslo Børs.

Totalt gir dette oss et bredt utvalg uavhengige variabler som danner ett godt grunnlag for analysene. Hvordan vi har behandlet verdiene kommer vi tilbake til i 4.1 Datatransformering.

4. Metode

I dette kapitlet vil vi gå inn på hvordan vi ønsker å få svar på problemstillingen vår. Vi har valgt å bruke maskinlæringsplattformen til Microsoft, kalt Azure Machine Learning. Grunnen til at vi har valgt denne programvaren er at det ikke kreves kraftig maskinvare lokalt siden den er skybasert. Dette gir oss gode muligheter for å kunne skalere modeller som blir krevende maskinelt uten at det krever investeringer i maskinvare og utstyr.

De fleste problemer som skal løses med maskinlæring følger samme trinnvise prosess:



Figur 4.1 - Maskinlæring standard teknikk

Som nevnt innledningsvis ønsker vi å finne ut om hvor godt ulike maskinlæringsteknikker kan brukes for å finne predikere endring på Oslo børs basert på variablene vi har valgt. Disse maskinlæringsteknikkene bruker ulike metoder for å predikere en avhengig variabel ut i fra en eller flere uavhengige variabler.

Vi har valgt å definere den avhengige variabelen som skal predikeres som en binær klassifiseringsvariabel – det vil si at den avhengige variabelen som skal defineres har to utfall. Helt konkret skal vi predikere om den avhengige variabelen blir å øke (at børsen går opp) eller

synke (at børsen går ned) ved Oslo Børs stengt tid ut i fra dagens endringer i de utvalgte variablene.

I de neste delkapitlene vil vi redegjøre for hvordan vi behandlet rådata for å bygge datasettet. Etter dette går vi inn på splitting av data og hvordan de ulike modellene blir generert. Deretter vil vi forklare hvordan man kan analysere resultatene og hvilke måleparametere vi har valgt å benytte. Til slutt vil vi si litt om reliabilitet, validitet og etiske aspekter tilknyttet designet vi har valgt.

4.1. Datatransformering

Datatransformering handler om å transformere rådata fra ulike kilder til et datasett som kan brukes til analyse. For å gjøre det lettere for oss å slå sammen datasettene har vi valgt å sette opp en lokal SQL database med rådata. Databasen har blitt brukt for å konvertere data gjennom bruk av programmeringsspråket T-SQL.

Delkapittelet starter med en diskusjon rundt hvordan vi har valgt å håndtere det at datasettene har ulike noteringstidspunkt, for å så gjennomgå konkret hvordan rådata er behandlet og satt sammen til et datasett som kan brukes i analyser. Til slutt er det en oppsummering med en oversikt over hvordan det ferdige datasettet ser ut.

4.1.1. Åpningstider på børsene

Som nevnt tidligere har vi i våre analyser valgt å benytte daglige observasjoner i perioden 2001-2017. Når vi jobber med daglige observasjoner kan det være problematisk at variablene våre ikke nødvendigvis er notert på samme tid. Alle variablene som vi har brukt i våre analyser ser på endring gjennom det siste døgnet fra foregående dag til inneværende dag hvor hver daglig observasjon.

For SSE Composite vil ikke dette være et problem for vår analyse da hele handletiden på disse børsene er før Oslo Børs åpner. Eventuelle effekter som disse børsene har på Oslo Børs vil komme ved åpningen av Oslo Børs og fanges opp av inneværende dag sin endring på denne børsen.

For de børsene som ligger i Europa er det små variasjoner i åpningstider og tidssoner. Dette betyr i teorien at det kan være endringer som oppstår etter Oslo Børs har stengt, men siden tidspunktene kun varierer med en til to timer har vi valgt å ikke gjøre tilpasninger for å ta hensyn til dette. Mesteparten av handletiden på de europeiske børsene er i samme handletid som Oslo Børs.

Angående S&P500 er det betydelig forskjell i åpningstid mot Oslo Børs. Børsene som handler aksjene som ligger i S&P 500 indeksen har en åpningstid i UTC mellom 14:30 og 21:00. Til sammenligning har Oslo Børs åpningstid mellom 08:00-15:30 i UTC tid. Det vil si at det kun er en time handel på Oslo Børs mot slutten av åpningstiden som er samtidig som amerikanske børser. Dette gjør at vi har tatt noen grep i forhold til hvilke data som vi har benyttet. Dette vil gjennomgås i neste delkapittel.

4.1.2. Oppbygging av datasett

Fra datainnsamlingen har en rekke CSV filer som inneholder daglige observasjoner av våre variabler. Disse er:

- En fil med daglige observasjoner av hovedindeksen på Oslo Børs hentet fra Oslo Børs.
- Åtte enkeltfiler for hver enkelt av de resterende børsindeksene fra Thompson Reuters.
- En fil for daglige observasjoner av Brent Spot hentet fra U.S. Energy Information Administration.
- En fil med daglige observasjoner på en rekke valuta hentet fra Norges Bank.
- En fil med daglige observasjonen på VIX indeksen hentet fra Yahoo Finance.

Filene har vi importert inn i en lokal database, og laget en database-tabell per fil. Hver enkelt observasjon av indeksene har kolonnene dato, sluttpris og åpningspris i tabellene. Valutaene og Brent har kolonnene dato og pris.

For indeks-, oljepris- og valuta-variablene våre har vi valgt å bruke endringsform. For å få disse på endringsform la vi inn nye kolonner i de respektive tabellene som skal brukes for å lagre prosentvis endring i verdien for siste døgn.

For hver av aksjeindeksene har vi kalkulert prosentvis endring som har blitt lagt til ($r_{i,t}$), hvor i er variabelens verdi og t er tid, utfra endring mellom sluttpris dagen på tabellraden (END_d) og sluttprisen på siste observasjon før inneværende dag (END_{d-1}). Formelen som er benyttet for å beregne endringen er:

$$r_{i,t} = \frac{END_d - END_{d-1}}{END_{d-1}} \quad (4.1)$$

Valutakursene har vi behandlet på lik måte, men her er det ikke flere verdier per dag. På valutakursene har vi benyttet verdien fra gitte dato (C_d) og siste observasjon før den gitte datoen

(C_{d-1}) for å komme frem til endringsverdien $(C_{i,t})$. Formelen vi har benyttet for å kalkulere valutakursendring:

$$C_{i,t} = \frac{(C_d) - (C_{d-1})}{(C_{d-1})} \quad (4.2)$$

Spotprisen på Brent er også kun en observasjon per dag og vi vil benytte en lik formel som valutaendringsformelen for å finne oljeprisens avkastning.

VIX, som er en prosentvis representasjon av forventede svingninger i markedet, har vi ikke konvertert til endringsform. Denne viser en prosentvis forventet volatilitet i markedet. Vi har heller ikke konvertert omsetning på Oslo Børs til endringsform da det er et kvantitativt mål på hvor mye likviditet som har vært i markedet den enkelte dagen.

Etter denne konverteringen har alle de respektive tabellene endringsverdiene for hver dag. For å sammenstille dette til et datasett tok vi utgangspunkt i OSEBX tabellen og slo den sammen med alle tabellene. Vi inkluderer kun de dagene hvor OSEBX har en observasjon. Datoer hvor Oslo Børs holder stengt (helger og helligdager) har vi ikke observasjoner. Grunnen til at vi velger å ikke inkludere dagene hvor børsen er stengt er på grunn av dette er den avhengige variabelen. Ved å sette inn en estimert verdi her som ofte gjøres i tidsserie-analyse med for eksempel et gjennomsnitt av observasjoner før og etter, vil metoden vår prøve å predikere disse utfallet av disse estimerte verdiene. Predikeringer som er gjort på bakgrunn av et datasett som har en stor andel fiktive verdier vil være lite relevant å bruke til analyse.

Det er gjort et unntak i forhold til sammenslåing av tabellene knyttet til S&P500-indeksen. På grunn av stor differanse i åpningstid mellom Oslo Børs og børsene i USA har vi valgt å benytte oss av to uavhengige variabler på denne indeksen. Den første variabelen vi er endringsverdien for observasjonen dagen før dagens observasjon. Dette vil da være på samme måte som det har blitt gjort med de andre børsene, men vi har lagt inn en dag forsinkelse (lag).

Den andre variabelen vi har valgt å benytte er endringen fra sluttprisen dagen før til åpningsprisen inneværende dag. Dette vil kunne bidra til å plukke opp eventuelle effekter en sterk eller svak åpning på børsene i USA vil ha på Oslo Børs den siste timen før stenging. Denne er kalkulert på følgende måte:

$$U_{i,t} = \frac{OPEN_d - END_{d-1}}{END_{d-1}} \quad (4.3)$$

Det er også valgt å inkludere foregående dags endring på Oslo Børs som uavhengig variabel i datasettet. Dette er lagt til på samme måte som foregående dags endring på S&P 500.

I den samme operasjonen som slår sammen tabellene har vi også lagt til to nye kolonner som skal med i analysen. Den første kolonnen er en binomialverdi for om hovedindeksen på Oslo Børs går opp eller ned. Denne skal brukes som vår avhengige variabel i datasettet. Denne er kalkulert fra dagens endring på OSEBX og har fått verdien 1 for oppgang og verdi 0 for nedgang. Vi har valgt å ekskludere de observasjonene vi har hvor OSEBX har vært uendret. Dette tilsvarer 22 av 4273 observasjoner i vårt datautvalg.

Den andre kolonnen som er lagt til er en kolonne for ukedag. Denne er kalkulert ut i fra dato-kolonnen gjennom funksjonen Datepart. Verdiene som har blitt lagt inn er verdien 1 for mandag til og med 5 for fredager. Vi har som sagt ingen observasjoner på helgedager da Oslo Børs er stengt.

I de tilfellene hvor vi mangler observasjoner på de uavhengige variablene har vi valgt å sette inn en verdi. For alle variablene våre på endringsform har vi satt endringen til 0 dersom det mangler en observasjon av variabelen på en enkelt dag. Vi har satt verdien til 0 siden det ikke er noen endring i variabelens verdi denne dagen, eksempelvis at den enkelte børsen er stengt den dagen. De dagene vi har manglet observasjon på VIX har vi valgt å sette inn verdien som var på den foregående dagen. Vi hadde også en enkelt dag hvor vi manglet omsetning på Oslo Børs, dette var 19.12.2017. Denne dagen har vi valgt å fjerne fra datasettet.

Tabellen nedenfor viser en oversikt over alle variablene i vårt datasett med tilhørende beskrivelse.

Navn	Datatype	Beskrivelse
DATE	Date	Dato for observasjonen
OSEBX	Binomialverdi	Avhengig variabel for om Oslo Børs gikk opp eller ned denne dagen. 1 for oppgang og 0 for nedgang.
DAY_OF_WEEK	Heltall	Verdi for hvilken dag i uken observasjonen gjelder. Alle observasjonene inneholder verdier mellom 1 (mandag) og 5 (fredag).
OSEBX_1	Desimaltall	Prosentvis endring i sluttpris fra to dager tilbake til gårdsdagen på Oslo børs hovedindeks.
BRENT	Desimaltall	Prosentvis endring i spotpris på Brent fra foregående dag til dagens pris.
AEX	Desimaltall	Prosentvis endring i sluttpris på AEX fra foregående dag til i dag.
SSE	Desimaltall	Prosentvis endring i sluttpris på SSE Comp fra foregående dag til i dag.
DAX	Desimaltall	Prosentvis endring i sluttpris på DAX fra foregående dag til i dag.
DK_OMX	Desimaltall	Prosentvis endring i sluttpris på OMX Copenhagen 20 fra foregående dag til i dag.
CAC40	Desimaltall	Prosentvis endring i sluttpris på CAC40 fra foregående dag til i dag.
FTSE100	Desimaltall	Prosentvis endring i sluttpris på FTSE 100 fra foregående dag til i dag.
SPCOMP_1	Desimaltall	Prosentvis endring i sluttpris fra to dager tilbake til gårdsdagen på S&P500.
SPCOMP_O	Desimaltall	Prosentvis endring i åpningspris fra foregående dag sin sluttpris på S&P500.
SE_OMX	Desimaltall	Prosentvis endring i sluttpris på OMX Stockholm 30 fra foregående dag til i dag.
USD	Desimaltall	Prosentvis endring i valutakurs NOK/USD fra foregående dag til i dag.
EUR	Desimaltall	Prosentvis endring i valutakurs NOK/EUR fra foregående dag til i dag.
GBP	Desimaltall	Prosentvis endring i valutakurs NOK/GBP fra foregående dag til i dag.
VIX	Desimaltall	Dagens sluttverdi på volatilitetsindeksen VIX.
OSEBX_T	Heltall	Daglig omsetning på Oslo Børs målt i Norske kroner.

Tabell 4.1 - Oversikt over inkluderte variabler

Tabell 4.2 inneholder deskriptiv statistikk fra vårt datasett. Minimumsverdien (Min) og maksimumsverdien (Max) angir laveste/høyeste verdi per variabel. For variablene på endringsform angir dette største nedgang/oppgang i perioden, mens for VIX og OSEBX_T angir det laveste/høyeste notering. SD er standardavviket til variablene.

Variabel	Min	Max	Gjennomsnitt	SD
OSEBX_1	-0,0995	0,1067	0,0004	0,0146
BRENT	-0,1804	0,1988	0,0005	0,0222
AEX	-0,0914	0,1055	0,0001	0,0145
SSE	-0,0884	0,0986	0,0002	0,0155
DAX	-0,0849	0,1140	0,0003	0,0151
DK_OMX	-0,1106	0,0996	0,0004	0,0128
CAC40	-0,0904	0,1118	0,0001	0,0147
FTSE100	-0,0885	0,0984	0,0001	0,0119
SPCOMP_1	-0,0903	0,1158	0,0002	0,0119
SPCOMP_O	-0,0284	0,0301	0,0000	0,0022
SE_OMX	-0,0842	0,1037	0,0002	0,0147
USD	-0,0575	0,0451	0,0000	0,0077
EUR	-0,0388	0,0398	0,0000	0,0048
GBP	-0,0559	0,0361	0,0000	0,0062
VIX	0,0923	0,8074	0,1984	0,0902
OSEBX_T	410 036 109	28 742 771 422	4 392 665 771	2 917 025 297

Tabell 4.2 - Deskriptiv statistikk fra datasettet

I Appendiks A er det lagt ved et utdrag fra vårt datasett.

4.1.3. Editering av Metadata

Før datasettet kan brukes til å generere modeller ved hjelp av maskinlæring, må man gi tilleggsinformasjon om variablene i datasettet. Dette gjøres gjennom å editere metadata. Metadata er informasjon om hva slags type data som en kolonne i datasettet inneholder. Ved å editere metadata til kolonnene tilegner man hver kolonne ekstra informasjon uten å redigere selve verdien. Dette er nødvendig i maskinlæring for at maskinlæringsteknikkene skal kunne håndtere dataen som ligger der på en logisk måte.

I vårt datasett har vi redigert metadataen på følgende måte:

- Kolonnen for Dato er kategorisert som en DateTime, slik at den blir tolket som et dato-felt.
- Alle kolonnene bortsett fra OSEBX er klassifisert som en feature i datasettet, slik at de klassifiseres som de uavhengige variablene i vårt datasett.

- Kolonnen for ukedag (Day_Of_Week) er klassifisert som kategorisk. Ved å klassifisere den som kategorisk vil verdien brukes for å gruppere data og ikke brukes i numeriske kalkulasjoner.
- Kolonnen OSEBX som inneholder 0 og 1 for om Oslo Børs skal gå opp eller ned er klassifisert som en Label. Dette forteller maskinlæringsalgoritmene at dette er den avhengige variabelen som skal kalkuleres ut i fra de uavhengige variablene. I tillegg siden det er en verdi som enten er sann eller usann for om at Oslo Børs skal gå opp eller ned, er den klassifisert som en kategorisk verdi med to utfall.

4.1.4. Normalisering av data for lineære modeller

Når vi genererer lineære modeller så vil vi få vekting på hver uavhengig variabel på hvor mye en endring i en verdi vil påvirke den avhengige variabelen. Ettersom vi har et datasett med variabler på ulik skala, er verdiene i datasettet svært forskjellig. Uten normalisering vil det være lite hensiktsmessig å sammenligne vektingen som modellene tilordner de ulike variablene. Eksempelvis har omsetningen på Oslo Børs verdi i milliardklassen, mens endringen på aksjeindeksene er på noen få prosentpoeng. Normalisering unngår dette problemet ved å generere nye verdier som opprettholder generell distribusjon og forhold mellom verdiene per variabel, samtidig som verdier holdes innenfor en felles skala for alle variablene som brukes i modellen. På vårt datasett har vi benyttet normaliseringsteknikken Z-score før vi genererte de lineære modellene.

Z-score (Z) kan fremstilles med følgende formel:

$$Z = \frac{x - \mu}{\sigma} \quad (4.4)$$

Hvor x er opprinnelig verdi i datasettet, μ er gjennomsnittet av alle verdiene x og σ er standardavviket til alle verdiene x . Ved å gjennomføre normalisering før man genererer modeller vil vektingen som genereres være mulig å bruke til å diskutere variablenes signifikans.

I Appendix B har vi lagt ved et utdrag fra vårt datasett etter normalisering.

4.2. Splitting av data

Når dataen har blitt konvertert er det neste steget å splitte dataen i to grupper. Den første gruppen av data skal brukes for å trene opp en modell ved bruk av maskinlæring, denne gruppen kalles treningsdata og det er her størsteparten av dataen skal legges inn. Grunnen til at denne gruppen skal være større enn den andre er fordi utvalget bør være stort for å kunne lage en mest

mulig representativ modell for datagrunnlaget. Ved å ha et lite utvalg av data i gruppen som skal brukes for å trene en modell vil det være større sannsynlighet for overtilpasning.

Den andre gruppen data (testdata) skal brukes for å teste modellen som har blitt generert ut i fra den første gruppen data. Her vil den resterende mengden data befinne seg. Vi går nærmere inn på hvordan denne dataen brukes i kapittelet om score- og evaluering av modell.

I vårt datasett var det totalt 4250 observasjoner hvorav 2300 er dager med oppgang og 1950 er dager med nedgang på Oslo Børs. Vi delte datasettet i to ved hjelp av stratifisert deling, som gir oss et tilfeldig utvalg med lik andel klassifisering 0 og 1 i henholdsvis trenings- og testdata. Vi har lagt 70 % av data i treningsdatasettet og de resterende 30 % lagt i testdatasettet. Treningsdatasettet inneholdt da 2975 observasjoner hvor 1610 av de var dager med oppgang på Oslo Børs og 1365 dager var dager med nedgang. I testdatasettet var det totalt 1275 observasjoner hvorav 690 var dager med oppgang og 585 var dager med nedgang.

Vi valgte å gjennomføre tre delinger med ulikt random seed (jfr. 2.2.5.5) slik at vi stod igjen med tre treningsdatasett med tilhørende testdatasett. Vi har valgt å gjøre dette for å kunne generere tre ulike modeller ut i fra treningsdata og teste de mot testdata, henholdsvis modell 1, 2 og 3. Det gir oss et bedre bilde av variansen til metodene, ettersom vi gjennomfører samme forsøk tre ganger med ulike datasett. Ved å generere tre ulike modeller vil vi også redusere sannsynligheten for at tilfeldigheter gjør at en metode er bedre enn en annen metode. Det er noe som kan oppstå hvis tilfeldigvis en enkelt deling passer best en metode. Ved å gjøre delingen tre ganger får vi ett bedre grunnlag for å diskutere hvilke(n) metode som egner seg best til å løse vårt problem.

4.3. Generering av modell

Som nevnt i kapittel 4.2 vil vi generere tre modeller per maskinlæringsmetode med ulike treningsdatasett. Når man genererer modeller ved hjelp av maskinlæringsmetode er det en rekke konfigureringer som må gjøres. Disse konfigurasjonene blir kalt hyperparametere. Målet med å konfigurere hyperparametere er å sette begrensninger til hver enkelt metode for å få en best mulig prediksjonsevne.

Hver maskinlæringsalgoritme har en rekke hyperparametere som kan settes for å øke ytelsen til modellene og/eller redusere tiden det tar å trene en modell ut i fra treningsdata. Appendix C er en oversikt over hvilke hyperparametere som er brukt per metode med tilhørende forklaring. For å få en mest mulig optimal sammensetning av hyperparametere har vi valgt å sette disse

igjennom hyperparameter tuning. Hyperparameter tuning genererer en rekke modeller iterativt med ulike tilfeldige utvalg verdier på hyperparameterene. Dette øker treningstiden betraktelig da hver metode vil generere mange modeller per datasett i stedet for kun en. For hvert treningsdatasett har vi benyttet den modellen som har blitt rangert høyest etter hyperparameter tuning. Dette er en prosess som kan kjøres ubegrenset antall ganger, men per treningsdatasett har vi satt denne til å kjøre en time. Siden det er stor variasjon i kompleksitet på de ulike metodene vil det være ulikt antall iterasjoner som er kjørt på hvert treningsdatasett. Hvor mange iterasjoner hver metode har kjørt på hvert treningsdatasett kommer vi tilbake til i hvert av underkapitlene i kapittel 5.

4.4. Evaluering av modell

Når maskinlæringsmodellene har blitt laget ut i fra treningsdatasettet, testes den på testdatasettet. Hver modell vil ut i fra de uavhengige variablene kalkulere resultatet på den avhengige variabelen. Det som blir kalkulert for hver eneste rad i testdatasettet er en sannsynlighetsverdi for at utfallet er 1 (positivt), i vårt tilfelle at børsen går opp, mellom 0 og 1. En verdi på 1 tilsvarer at modellen mener at det er 100% sannsynlighet for at børsen går opp den angitte dagen. For å evaluere resultatene av kalkuleringene tar vi den inn i en «confusion matrix» som deler resultatene av kalkuleringen i fire kategorier.

	Faktisk	
Predikert	TP	FP
	FN	TN

Figur 4.2 - Prediksjonskategorier

Disse kategoriene er:

- True positive (TP) – verdier som ble kalkulert sann, og er faktisk sann. I vårt tilfelle betyr dette at modellen predikerer oppgang på OSEBX når det er faktisk oppgang.
- False positive (FP) – verdier som ble kalkulert sann, men er faktisk usann (Type II Error). I vårt tilfelle betyr dette at modellen predikerer oppgang på OSEBX når det er faktisk nedgang.
- True negative (TN) – verdier som ble kalkulert usann, og er faktisk usann. I vårt tilfelle betyr dette at modellen predikerer nedgang på OSEBX når det faktisk er nedgang.
- False negative (FN) – verdier som ble kalkulert usann, men er faktisk sann (Type I Error). I vårt tilfelle betyr dette at modellen predikerer nedgang på OSEBX når det faktisk er oppgang.

Det finnes flere ulike måter å analysere disse resultatene på som vi skal gå nærmere inn på nå. Avhengig av type problem vil det være mer hensiktsmessig å legge vekt på en måleparameter foran en annen.

4.4.1. Accuracy

Accuracy måler hvor stor andel av kalkuleringene som er korrekte. Det er en verdi mellom 0 og 1, hvor en større verdi indikerer bedre Accuracy.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (4.5)$$

4.4.2. Precision

Precision måler hvor stor andel av kalkuleringene som er positiv som faktisk er positiv. Dette er en verdi mellom 0 og 1, hvor en større verdi indikerer bedre presisjon.

$$Precision = \frac{TP}{TP + FP} \quad (4.6)$$

4.4.3. Recall

Recall, også kalt true positive rate, måler hvor stor andel som faktisk er positiv og er kalkulert som positiv. Dette er en verdi mellom 0 og 1, hvor en større verdi indikerer bedre Recall.

$$Recall = \frac{TP}{TP + FN} \quad (4.7)$$

4.4.4. F-Score

F-Score er en måleparameter som tar et vektet gjennomsnitt av Precision og Recall. Denne måleparameteren tar med andre ord både hensyn til false positives og false negatives.

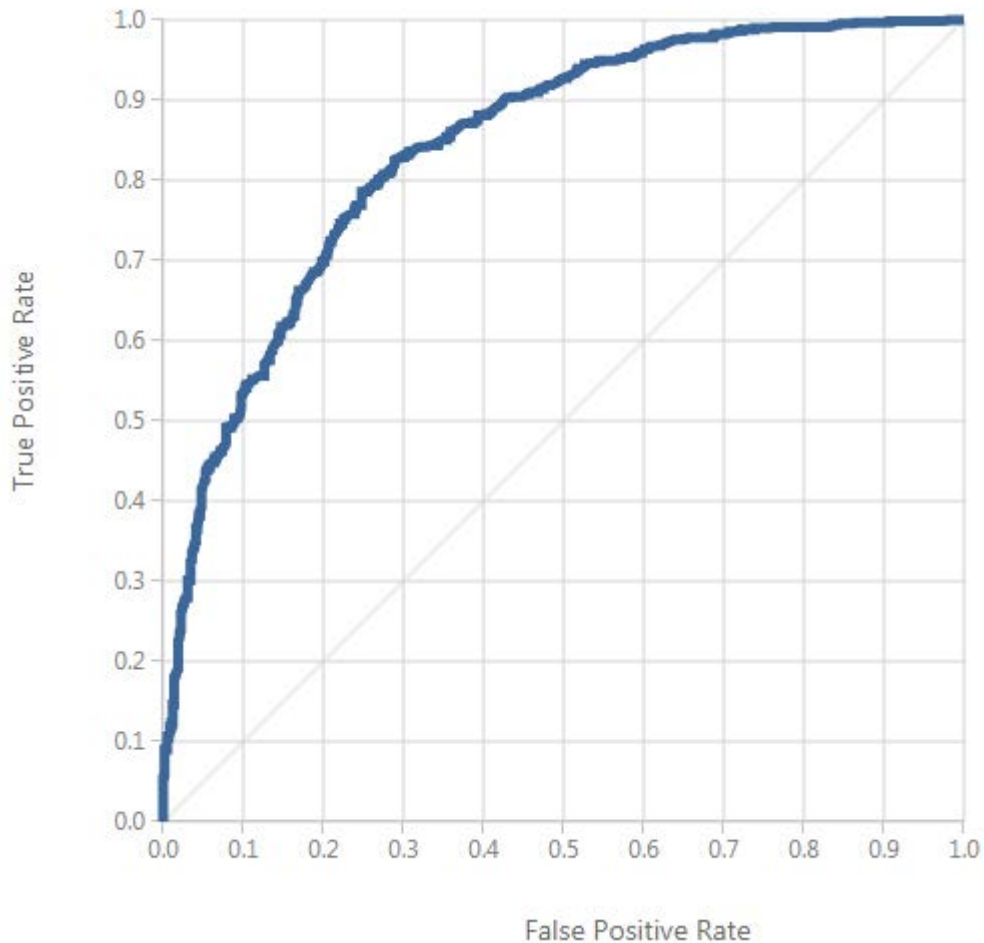
$$F - Score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4.8)$$

4.4.5. Area under the curve (AUC)

En annen måte å analysere ytelsen til en modell på er ved bruk av en «Receiver Operating Characteristic» (ROC) kurve, herunder måling av «Area Under the Curve» (AUC) verdi. En ROC kurve gir oss en grafisk fremstilling på hvor god klassifiseringsytelse en modell har. ROC kurven kan vi generere ved å lage en «confusion matrix» med ulike terskelverdier for hvilke prosent sannsynlighet man skal godta for å kalkulere et sant utfall i modellen. For hver av de ulike terskelverdiene regner man ut true positive rate (TPR) og false positive rate (FPR). Formelen for TPR (også kalt Recall) er nevnt tidligere, mens formelen for FPR er:

$$FPR = \frac{TN}{TN + FP} \quad (4.9)$$

Etter utregningene kan vi tegne opp en ROC kurve. Dette gjøres ved at vi for hver av terskelverdiene, eksempelvis fra 0 til 1 med 0,01 i intervall – plotter inn verdien for FPR i X-aksen og TPR i Y-aksen for å så tegne en kurve igjennom punktene.



Figur 4.3 - Eksempel på ROC-kurve

Figur 4.3 viser ROC-kurven fra ett av våre forsøk. Ideelt sett klatrer en ROC-kurve raskt opp mot øverste venstre hjørne, noe som betyr at modellen har kalkulert riktig de fleste tilfellene. Dess nærmere denne kurven er øvre venstre hjørne, jo bedre er klassifiseringsytelsen. For å oppnå dette må modellen maksimere den sanne positive frekvensen og minimere den falske positive frekvensen. Dersom kurven som ligger nær diagonalen, tyder det på en modell som gjør klassifiseringer som er nær tilfeldig gjetning.

Området under kurven defineres som «Area Under the Curve» (AUC). AUC kalkuleres til en konkret verdi mellom 0 og 1 som definerer hvor stor andel av det todimensjonale rommet som ligger under grafen. Her vil en verdi på 0,5 tilsvare en tilfeldig klassifikator, mens en verdi på 1 vil være en perfekt klassifikator.

4.4.6. Logarithmic Loss

En annen måleparameter som vi benytter for å vurdere modeller er Logarithmic Loss (log loss). Log loss kan kalkuleres på følgende vis:

$$LogLoss = -\frac{1}{N} \sum_{n=0}^{N-1} \sum_{k=0}^{K-1} y_{n,k} \log[p_{n,k}] \quad (4.10)$$

Hvor

- N er antallet predikeringer
- K er antallet mulige klassifiseringer. Siden vi jobber med binære klassifiseringer dette antallet 2.
- $y_{n,k}$ er en binær indikator om en klassifisering k er korrekt klassifisering for observasjon n , i vårt tilfelle vil denne ha verdi 1 dersom predikeringen er korrekt, og 0 dersom predikeringen er feil.
- $p_{n,k}$ er sannsynligheten assosiert med prediksjon n av en klassifisering k .

Siden vi jobber med binær klassifisering kan log loss forenkles til:

$$LogLoss = -\frac{1}{N} \sum_{n=0}^{N-1} [y_n \log p_n + (1 - y_n) \log(1 - p_n)] \quad (4.11)$$

En perfekt klassifiseringsmodell vil ha en log loss lik 0. Dess lenger log loss verdien er i fra 0, dess dårligere er modellen på å predikere. En modell som klassifiserer 100 % korrekt, men med 50 % sannsynlighet vil få en log loss på 0,693. En log loss $> 0,693$ er med andre ord ikke en akseptabel verdi. Log loss straffer klassifiseringer som er predikert med en klassifisering k med stor sannsynlighet som er predikert feil.

Log loss er en måleparameter som er spesielt interessant i en økonomisk kontekst. Dette er fordi den straffer spesielt de tilfellene hvor det blir predikert med stor sannsynlighet og predikeringen er feil. Hvis vi setter dette inn i vår kontekst og eksempelvis skulle ha brukt disse modellene for å ta avgjørelser på om det skal investeres på OSEBX. Da vil en modell som, med 99 % sannsynlighet, predikerer oppgang på en dag med nedgang, være en dårligere modell enn en som predikerer oppgang med 55 % sannsynlighet. I dette eksempelet vil Accuracy, Precision, Recall og F-Score være like på begge modellene, mens log loss vil være mye høyere på den første. Ved å inkludere log loss, så får vi ett bedre bilde av ytelsen til modellene.

4.5. Reliabilitet

I forskningen er man avhengig av nøyaktige data for å kunne gjennomføre gode analyser. En målestokk på dette kalles reliabilitet. Det er viktig at undersøkelsen skal kunne gjenskapes av andre forskere, og det er viktig at dersom de brukes samme metode og samme datasett så skal de komme frem til samme resultat.

Vi benytter oss av sekundærdata i form av gitte verdier fra blant annet SSB, Oslo Børs, Thomson Reuters og Norges Bank som er anerkjente institusjoner som ofte benyttes i forskningssammenheng. Vi føler oss derfor trygge på at datainnsamlingen vår oppfyller krav til reliabilitet. Videre må en spørre seg hvorvidt databehandlingen vår vil endre reliabiliteten til datasettet. Vi har beskrevet godt hvordan vi har behandlet rådata for å lage endelig datasett. Vi er trygge på at metoden vi har valgt ikke har manipulert dataen slik at funnene er blitt påvirket. Totalt sett så føler vi at datainnsamling og databehandling oppfyller kravene til reliabilitet.

4.6. Validitet

Validitet sier noe om i hvilken grad man evner å svare på det man ønsker. Er variablene valgt de riktige for å gi svar på problemstillingen? Validitet skilles i to kategorier, intern og ekstern validitet.

4.6.1. Intern validitet

For å finne svar på problemstillingen så må man gjøre avgrensninger i forhold til datainnsamling. Man vil aldri kunne undersøke alle parametere som påvirker et fenomen og således kan man si at det er umulig å oppnå full validitet. Derfor handler det om å velge de parametere som man enten selv mener, eller kan bevise at henger sammen med det man forsøker å finne ut av. Dette kalles kausalsammenhenger (Johannessen et. Al, s. 324, 2009) I vår studie ønsker vi å svare på om vi kan forutse endringer i det norske aksjemarkedet ved hjelp av en rekke økonomiske faktorer og internasjonale aksjemarkeder. Det vil naturlig nok være mange flere faktorer enn de vi tar med som vil påvirke børsen, slik som for eksempel nyheter, kvartalsrapporter mm., men som nevnt i kapittelet om motivasjon så finnes det eksempler på annen forskning med lik problemstilling som har benyttet lignende variabler som vi har valgt ut. Vi har i tillegg ett bredere utvalg variabler enn tidligere leverte masteroppgaver med lignende problemstilling. Det gjør at vi trolig har økt intern validitet sammenlignet disse.

4.6.2. Ekstern validitet

Johannessen et. al (s. 411, 2011) skriver at ekstern validitet stiller «spørsmål om i hvilken grad resultater fra en undersøkelse kan overføres i rom og tid». I hvilken grad kan funnene i vår studie brukes til å forklare hendelser i andre tidsperioder og aksjemarkeder?

Vi mener vår eksterne validitet er nokså god. Variablene vi har valgt representerer godt det vi ønsker å gi svar på. Vi føler at det OSEBX er en god indikator på det norske aksjemarkedet ettersom den omfatter en stor andel av de likvide aksjene i Norge og den viser en sterk korrelasjon til OSEAX. Vi føler ikke at vi kunne brukt samme uavhengige variabler for å gi svar på børsbevegelser i andre land, da måtte man nok tilpasset de uavhengige variablene til det gitte aksjemarkedet. Vi mener med bakgrunn i drøftingen ovenfor at vår eksterne validitet er tilfredsstillende.

4.7. Forskningsetikk

Et forskningsprosjekt består av mange faktorer som man må ta hensyn til. I tillegg til blant annet validitet og reliabilitet som nevnt over, så må man ta stilling til hvorvidt prosjektet man ønsker å gjennomføre er lovlig fra et juridisk standpunkt, og man må også reflektere over hvilke etiske problemstillinger man kan møte. Mens man ofte fort kan definere hvorvidt et prosjekt er juridisk forsvarlig, så kan etiske problemstillinger være mer intrikate, og man må evne å sette seg inn i hvordan konsekvenser forskningen kan få for andre.

I vårt forskningsprosjekt har vi ikke brukt informanter eller respondenter, men kun sekundærdata. Ettersom vi har gjennomført et forskningsprosjekt med kun sekundærdata, så gir det ikke like mange etiske utfordringer som man potensielt kan møte i en kvalitativ studie med intervjuer. De nasjonale forskningsetiske komiteene har produsert forskningsetiske retningslinjer som gir en god indikasjon på hvilke faktorer man må tenke over (www.etikkom.no, 2018). For vår del blir derfor de etiske aspektene knyttet til at vi er nøye i databehandlingen, følger god henvisningsskikk, og at vi gjør datamaterialet tilgjengelig for de som skulle ønske å benytte det til etterprøving eller videre forskning.

Videre er det viktig at vi rapporterer de funn vi finner, og ikke manipulerer data for å fremprovosere funn, eller rapporterer feil for å kunne vise til resultat.

5. Resultat og analyse

I dette kapitlet vil vi gå igjennom resultatene de ulike metodene har gitt oss. Vi vil først gjennomgå resultatene fra de lineære metodene og sammenligne de mot hverandre. Vi vil også sammenligne vektingen til de uavhengige variablene når det gjelder de lineære modellene. Deretter vil vi se på resultater fra de ikke-lineære metodene. Til slutt vil vi se på hvordan metodene scorer på måleparameterne vi har valgt å benytte, og sammenligne de mot hverandre.

5.1. Lineære metoder

I dette delkapitlet vil vi gjennomgå resultatene fra de tre lineære metodene vi har benyttet. Vi vil først gå inn på resultater for hver av de tre modellene, før vi kommenterer de ulike funnene mot hverandre.

Som nevnt tidligere vil de lineære modellene gi oss ulik vekting for de uavhengige variablene ut fra hvor mye den genererte modellen mener de påvirker den avhengige variabelen. Vektingene kan være enten positive eller negative ut fra hvordan modellene mener de påvirker utviklingen på OSEBX.

De uavhengige variablene er som nevnt i kapittel 4.1.4 normalisert ved å benytte Z-score. Når vektene er utregnet så har vi kalkulert prosentvis vekting på hver av de kvantitative variablene basert på absoluttverdien til vekten. Deretter har vi regnet gjennomsnittlig prosentvis vekting av de kvantitative variablene til de tre modellene.

Vi har valgt å trekke ut ukedagene i en egen tabell på hvert av delkapitlene. Siden ukedag er en kvalitativ verdi og ikke en kvantitativ verdi vil det være et ledd per kategori k av denne variabelen i de lineære modellene. Hvert ledd vil ha ulik vekt som ganges med en variabel som vil ha verdien 0 på de dagene den ikke er lik klassifisering k , og verdi 1 på den dagen ukedagen er lik klassifisering k . Eksempelvis vil variabelen mandag ha verdi 1 når utfallet på en mandag skal predikeres, mens verdien for de øvrige ukedagsvariablene er 0.

Ukedag benytter samme skala som tabellene som viser de kvantitative uavhengige variablene, slik at summen for de to tabellene (vekting av kvantitative variabler og ukedag) til sammen utgjør 100 %.

5.1.1. Bayes Point Machine

Bayes Point Machine er den eneste av maskinlæringsteknikkene som ikke støtter hyperparameter tuning. Dette er på grunn av at den eneste hyperparameteren man kan justere på denne algoritmen er antall iterasjoner den skal bruke for å konvergere på den optimale modellen. Vi satt opp hver av modellene som skulle bli generert til å kjøre 4000 iterasjoner før den konvergerer på den optimale gjennomsnittlige klassifiseringsmodellen.

5.1.1.1. Vekting av variabler

Modellene som ble skapt ga oss følgende vekt på de kvantitative uavhengige variablene:

Variabel	Modell 1		Modell 2		Modell 3		Gjennomsnitt
	Vekt	Prosent	Vekt	Prosent	Vekt	Prosent	Prosent
SWE_OMX	0,353	12,72 %	0,305	10,34 %	0,399	13,82 %	12,29 %
BRENT	0,348	12,54 %	0,338	11,45 %	0,339	11,74 %	11,91 %
FTSE100	0,370	13,33 %	0,313	10,61 %	0,323	11,18 %	11,71 %
DK_OMX	0,298	10,74 %	0,291	9,86 %	0,326	11,29 %	10,63 %
AEX	0,159	5,73 %	0,261	8,84 %	0,225	7,79 %	7,46 %
SPCOMP_O	0,237	8,54 %	0,215	7,29 %	0,126	4,36 %	6,73 %
CAC40	0,153	5,51 %	0,227	7,69 %	0,148	5,12 %	6,11 %
SPCOMP_1	0,163	5,87 %	0,158	5,35 %	0,178	6,16 %	5,80 %
OSEBX_1	-0,091	3,28 %	-0,113	3,83 %	-0,137	4,74 %	3,95 %
GBP	0,059	2,13 %	0,117	3,96 %	0,093	3,22 %	3,10 %
EUR	-0,077	2,77 %	-0,096	3,25 %	-0,090	3,12 %	3,05 %
DAX	-0,057	2,05 %	-0,092	3,12 %	-0,053	1,84 %	2,34 %
BIAS	0,047	1,69 %	0,046	1,56 %	0,048	1,66 %	1,64 %
SSE	0,012	0,43 %	0,017	0,58 %	0,077	2,67 %	1,22 %
USD	0,001	0,04 %	-0,053	1,80 %	0,029	1,00 %	0,95 %
VIX	0,006	0,22 %	0,004	0,14 %	0,050	1,73 %	0,69 %
OSEBX_T	0,005	0,18 %	0,025	0,85 %	-0,002	0,07 %	0,37 %
DATE	0,000	0,00 %	0,000	0,00 %	0,000	0,00 %	0,00 %

Tabell 5.1 - Vekting variabler BPM

Vi ser at børsene i Skandinavia samt Storbritannia er høyere vektet de resterende børsene, med positivt fortegn. Videre ser vi at alle børsene bortsett fra Tyskland er positivt vektet. Det tyder på at modellene mener at en oppgang på DAX påvirker OSEBX negativt. I tillegg ser vi at oljeprisen blir tillagt stor positiv vekt i alle modellene. Når vi ser på valutatallene så er det Euro og Pund som er de viktigste valutaen i modellen. Euro er tillagt negativ vekt, mens Pund er positivt vektet. Dollar ser ut til å være mindre signifikant. Volatilitetsindeksen og omsetning på Oslo Børs har også relativt liten påvirkning på modellene. Generelt sett så er det liten differanse i vektingen på de tre modellene, men vi ser at AEX, S&P500s åpningspris og CAC40 har litt større variasjon enn resten.

Tabell 5.2 viser vekting av ukedagene.

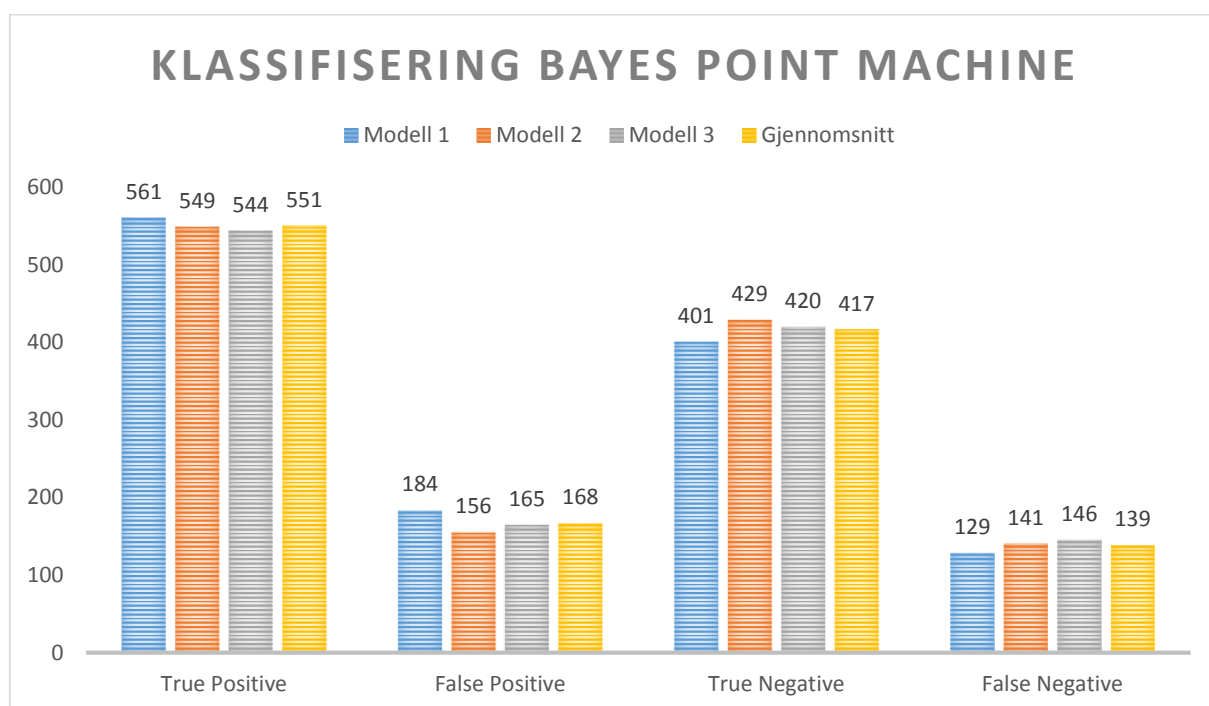
Ukedag	Modell 1		Modell 2		Modell 3		Gjennomsnitt Prosent
	Vekt	Prosent	Vekt	Prosent	Vekt	Prosent	
Torsdag	0,084	3,03 %	0,106	3,59 %	0,120	4,16 %	3,59 %
Onsdag	0,086	3,10 %	0,075	2,54 %	0,060	2,08 %	2,57 %
Tirsdag	0,102	3,68 %	0,074	2,51 %	0,024	0,83 %	2,34 %
Fredag	0,018	0,65 %	-0,018	0,61 %	0,034	1,18 %	0,81 %
Mandag	-0,049	1,77 %	-0,007	0,24 %	0,007	0,24 %	0,75 %

Tabell 5.2 - Vekting ukedag BPM

Modellene vekter generelt sett torsdag høyest av alle ukedagene, og mener dermed at ukedagseffekten er størst på denne dagen. Tirsdag og Onsdag er også en del høyere vektet enn fredag og mandag. Disse tre dagene er positivt vektet, mens mandag har en negativ verdi for modell 1 og 2 og fredag har det for modell 1. Dette tyder på at modellene som har vektet mandag og fredag negativt mener at det er en negativ ukedagseffekt tilknyttet disse dagene.

5.1.1.2. Prediksjonsytelse

Etter at modellene ble generert ut fra treningsdata så ble testdata kjørt gjennom modellene. Det ga oss følgende resultater:



Figur 5.1 - Klassifiseringer BPM

De tre modellene presterer nokså jevnt når de skal klassifisere oppgang (positive). De klarte i gjennomsnitt å predikere 551 av 690 dager med oppgang. Resultatene fra klassifiseringene gir

oss grunnlaget for å regne ut måleparameterne for modellene, som presenteres i Tabell 5.3, sammen med Log loss-verdiene.

		Accuracy	Precision	Recall	F-Score	AUC	Log Loss
Modell 1	Trening	0,772	0,777	0,813	0,795	0,858	0,472
	Test	0,755	0,753	0,813	0,782	0,842	0,484
	Differanse	-0,018	-0,024	0,000	-0,013	-0,016	0,013
Modell 2	Trening	0,771	0,771	0,820	0,795	0,859	0,466
	Test	0,767	0,779	0,796	0,787	0,841	0,497
	Differanse	-0,004	0,008	-0,025	-0,008	-0,018	0,031
Modell 3	Trening	0,770	0,776	0,809	0,792	0,860	0,461
	Test	0,756	0,767	0,788	0,778	0,835	0,509
	Differanse	-0,014	-0,008	-0,021	-0,014	-0,025	0,047
Gjennomsnitt	Trening	0,771	0,774	0,814	0,794	0,859	0,466
	Test	0,759	0,766	0,799	0,782	0,839	0,497
	Differanse	-0,012	-0,008	-0,015	-0,012	-0,020	0,030

Tabell 5.3 - Ytelse BPM

Generelt sett så er det liten differanse mellom trening og testdata. Det kan tyde på at modellene generaliserer relativt godt på testdata. Alle de tre modellene predikerer korrekt (Accuracy) i ca. 76 % av tilfellene. Ettersom Recall generelt er høyere enn Accuracy så tyder det på at modellene er bedre på å predikere oppgang enn nedgang på OSEBX.

5.1.2. Logistic Regression

For Logistic Regression forsøkte vi 10.000 ulike konfigurasjoner av hyperparametere på hvert treningsdatasett. Med andre ord ble det generert 10.000 ulike modeller per treningsdatasett. Vi har benyttet den optimale modellen fra hvert treningsdatasett. De tre optimale modellene ble deretter testet mot testdata.

5.1.2.1. Vekting av variabler

Modellene som ble skapt ga oss følgende vekt på de ulike variablene:

Variabel	Modell 1		Modell 2		Modell 3		Gjennomsnitt
	Vekt	Prosent	Vekt	Prosent	Vekt	Prosent	Prosent
BIAS	-21,632	29,75 %	-22,437	29,75 %	-14,362	27,95 %	29,15 %
BRENT	7,224	9,94 %	8,507	11,28 %	5,081	9,90 %	10,37 %
DK_OMX	7,321	10,07 %	7,683	10,19 %	5,403	10,51 %	10,26 %
SWE_OMX	7,393	10,17 %	6,685	8,86 %	5,480	10,66 %	9,90 %
FTSE100	8,145	11,20 %	6,481	8,59 %	4,421	8,60 %	9,46 %
AEX	4,299	5,91 %	5,802	7,69 %	3,831	7,45 %	7,02 %
CAC40	3,454	4,75 %	4,915	6,52 %	3,737	7,27 %	6,18 %
SPCOMP_O	5,252	7,22 %	5,607	7,43 %	0,525	1,02 %	5,22 %
SPCOMP_1	3,293	4,53 %	2,672	3,54 %	2,254	4,39 %	4,15 %
OSEBX_1	-1,494	2,06 %	-1,253	1,66 %	-0,823	1,60 %	1,77 %
DAX	0,601	0,83 %	0,000	0,00 %	2,155	4,19 %	1,67 %
EUR	-0,903	1,24 %	-1,067	1,42 %	-0,985	1,92 %	1,53 %
GBP	0,364	0,50 %	0,970	1,29 %	0,835	1,62 %	1,14 %
SSE	0,079	0,11 %	0,094	0,12 %	0,796	1,55 %	0,59 %
DATE	-0,486	0,67 %	-0,282	0,37 %	-0,278	0,54 %	0,53 %
USD	-0,024	0,03 %	-0,400	0,53 %	-0,094	0,18 %	0,25 %
VIX	-0,337	0,46 %	-0,009	0,01 %	-0,114	0,22 %	0,23 %
OSEBX_T	0,000	0,00 %	0,264	0,35 %	0,033	0,07 %	0,14 %

Tabell 5.4 - Vekting variabler Logistic Regression

BIAS (skjevhet) tillegges stor negativ vekt, mens de fleste andre verdiene er positive. Skjevhet er som nevnt ett feilledd som brukes for å korrigere modellen. Det tyder på at modellene kan være undertilpasset.

For øvrig viser resultatene at oljeprisen er en variabel som blir vektet tungt i modellene. Videre så ser man at våre nærmeste naboer, Danmark, Sverige og Storbritannia, er de børsene som blir vektet tyngst i prediksjonsmodellene. Valutakursene generelt er mindre signifikant enn børsene, med unntak av Kina. Det samme gjelder VIX og omsetning på Oslo Børs. Generelt så viser resultatene at en økning på børsvariablene tyder på økning på Oslo Børs, ettersom alle børsene

har positiv vekt. De som viser negativ vekt er gårsdagens endring på Oslo Børs, Euro, Dollar og VIX.

Det er til dels store forskjeller mellom hver av de tre ulike modellene, for eksempel vektet åpningsprisen på børsen i USA nokså likt i modell 1 og 2, mens modell 3 tillegger den en mye lavere vekt.

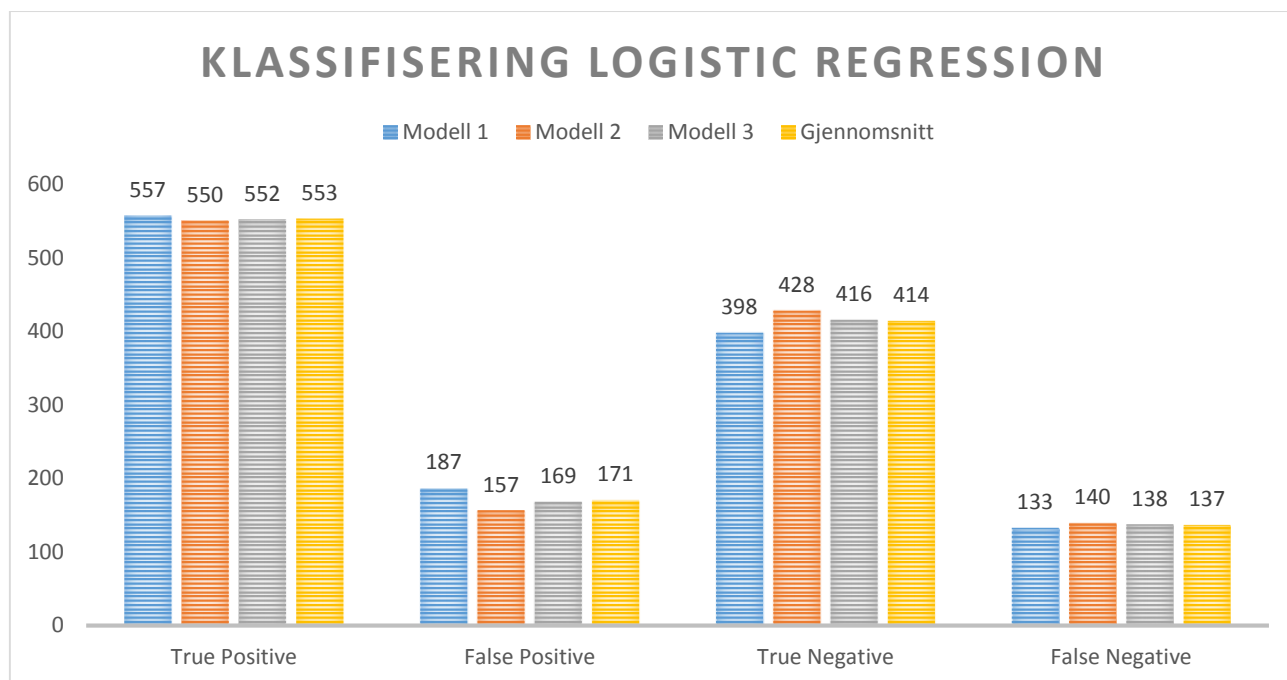
Variabel	Modell 1		Modell 2		Modell 3		Gjennomsnitt
	Vekt	Prosent	Vekt	Prosent	Vekt	Prosent	
Torsdag	0,097	0,13 %	0,118	0,16 %	0,124	0,24 %	0,18 %
Onsdag	0,104	0,14 %	0,060	0,08 %	-0,018	0,04 %	0,09 %
Mandag	-0,115	0,16 %	-0,056	0,07 %	0,000	0,00 %	0,08 %
Tirsdag	0,097	0,13 %	-0,001	0,00 %	-0,020	0,04 %	0,06 %
Fredag	0,000	0,00 %	-0,066	0,09 %	-0,020	0,04 %	0,04 %

Tabell 5.5 - Vekting ukedag Logistic Regression

Når vi ser på hvordan ukedag vektet i de ulike modellene så ser vi at torsdag igjen er tillagt størst vekt, med positivt fortegn. Mandag og fredag er tillagt negativ eller ingen vekt i de tre modellene. I modell 3 er torsdag markant tyngre vektet enn de andre ukedagene.

5.1.2.2. Prediksjonsytelse

Etter at modellene ble generert ut fra treningsdata så ble de respektive testdatasettene kjørt gjennom hver enkelt modell. Det ga oss følgende testresultater:



Figur 5.2 - Klassifiseringer Logistic Regression

Som vi ser så har Logistic Regression gjennomsnittlig klassifisert oppgang i 553 av de 690 dagene med oppgang. LR har en marginalt høyere andel FP-klassifiseringer enn BPM, altså predikert oppgang i tilfeller hvor datasettet viser at det faktisk var nedgang. Klassifiseringene ga oss grunnlaget for å regne ut følgende måleparameterne:

		Accuracy	Precision	Recall	F-Score	AUC	Log Loss
Modell 1	Trening	0,784	0,789	0,820	0,804	0,859	0,467
	Test	0,749	0,749	0,807	0,777	0,841	0,485
	Differanse	-0,035	-0,040	-0,013	-0,027	-0,018	0,018
Modell 2	Trening	0,775	0,776	0,820	0,798	0,859	0,464
	Test	0,767	0,778	0,797	0,787	0,842	0,490
	Differanse	-0,008	0,001	-0,023	-0,010	-0,017	0,026
Modell 3	Trening	0,774	0,771	0,828	0,799	0,858	0,471
	Test	0,759	0,766	0,800	0,782	0,837	0,503
	Differanse	-0,015	-0,006	-0,028	-0,016	-0,022	0,032
Gjennomsnitt	Trening	0,777	0,779	0,823	0,800	0,859	0,467
	Test	0,758	0,764	0,801	0,782	0,840	0,493
	Differanse	-0,019	-0,015	-0,021	-0,018	-0,019	0,025

Tabell 5.6 - Ytelse Logistic Regression

Generelt sett så er det også her liten differanse mellom trenings- og testdata, om enn noe dårligere enn BPM, som tyder på at modellene er marginalt dårligere til å generalisere. Med hensyn på testdata så er Accuracy nokså lik BPM, prediksjonene er korrekt i ca. 76 % av tilfellene. LR ser også ut til å være bedre til å predikere oppgang enn nedgang på OSEBX, siden Recall generelt er høyere enn Accuracy.

5.1.3. Support vector machine

For SVM forsøkte vi 1000 ulike konfigurasjoner av hyperparametere på hvert treningsdatasett. Med andre ord ble det generert 1000 ulike modeller per treningsdatasett. Vi har benyttet den optimale modellen fra hvert treningsdatasett. De tre optimale modellene ble deretter testet mot testdata.

5.1.3.1. Vekting av variabler

Modellene som ble skapt ga oss følgende vekt på de ulike variablene:

Variabel	Modell 1		Modell 2		Modell 3		Gjennomsnitt
	Vekt	Prosent	Vekt	Prosent	Vekt	Prosent	Prosent
BIAS	-9,903	26,06 %	-8,084	24,67 %	-7,389	25,35 %	25,36 %
SWE_OMX	3,801	10,00 %	3,229	9,85 %	2,790	9,57 %	9,82 %
CAC40	2,941	7,74 %	2,955	9,01 %	2,394	8,20 %	8,32 %
DK_OMX	3,015	7,93 %	2,575	7,86 %	2,669	9,16 %	8,32 %
FTSE100	3,027	7,97 %	2,540	7,75 %	2,368	8,12 %	7,95 %
AEX	2,951	7,77 %	2,648	8,08 %	2,297	7,88 %	7,91 %
BRENT	2,958	7,79 %	2,242	6,84 %	2,083	7,15 %	7,26 %
DAX	2,400	6,32 %	2,410	7,36 %	2,165	7,43 %	7,03 %
SPCOMP_1	1,447	3,81 %	0,682	2,08 %	0,972	3,34 %	3,08 %
VIX	-0,820	2,16 %	-0,903	2,76 %	-0,659	2,26 %	2,39 %
SPCOMP_O	1,149	3,02 %	0,742	2,26 %	0,520	1,78 %	2,36 %
EUR	-0,757	1,99 %	-0,736	2,25 %	-0,587	2,01 %	2,08 %
SSE	0,566	1,49 %	0,434	1,32 %	0,737	2,53 %	1,78 %
OSEBX_1	-0,569	1,50 %	-0,504	1,54 %	-0,550	1,89 %	1,64 %
USD	-0,458	1,21 %	-0,702	2,14 %	-0,223	0,76 %	1,37 %
DATE	-0,486	1,28 %	-0,282	0,86 %	-0,255	0,88 %	1,00 %
GBP	-0,150	0,39 %	-0,015	0,05 %	0,131	0,45 %	0,30 %
OSEBX_T	0,009	0,02 %	0,226	0,69 %	0,022	0,08 %	0,26 %

Tabell 5.7 - Vekting variabler SVM

I likhet med Logistic Regression-modellene så er BIAS(skjevhet) vektet høyt også i SVM-modellene, med negativt fortegn. Det kan igjen tyde på at modellene er undertilpasset. Vi ser at børsene er generelt høyt vektet. Igjen er våre nærmeste naboer blant de børsene som tillegges størst vekt, men SVM tillegger forholdsmessig større vekt på Frankrike og Nederland. Vi ser også at Tyskland tillegges mye større vekt i SVM-modellene enn BPM og LR. Volatilitetsindeksen er også tillagt betydelig større vekt enn i BPM og LR. Generelt sett så er de tre ulike modellene jevnere vektet per variabel enn BPM og LR.

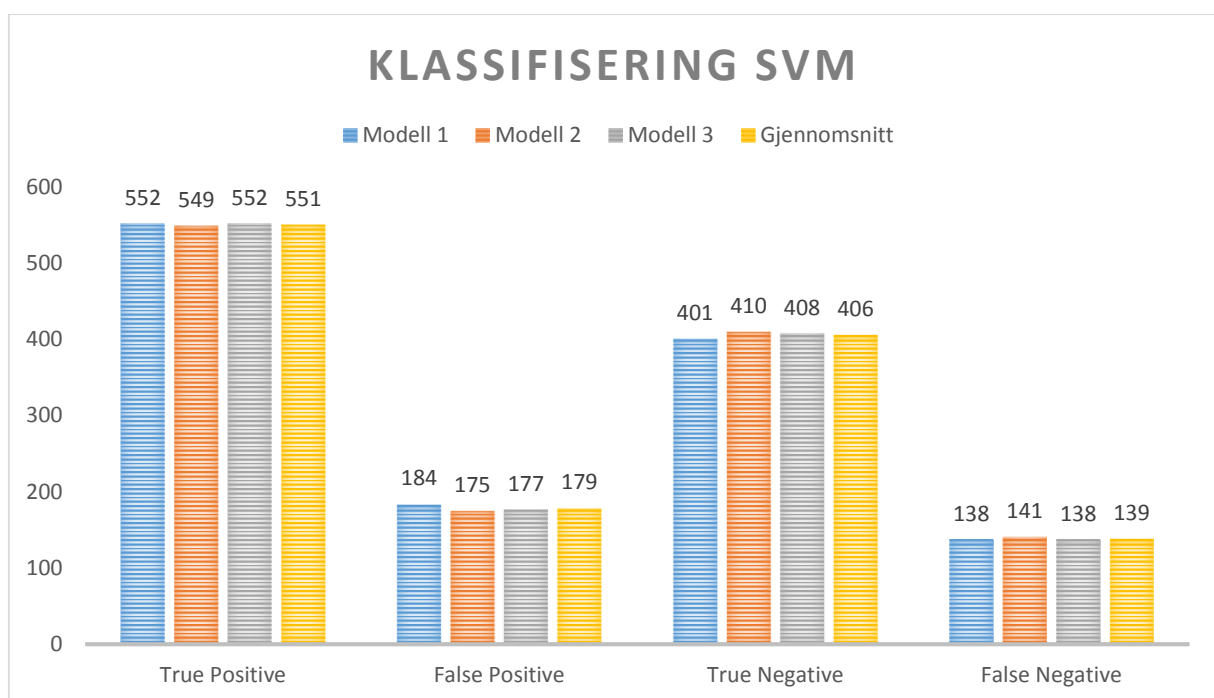
Ukedag	Modell 1		Modell 2		Modell 3		Gjennomsnitt Prosent
	Vekt	Prosent	Vekt	Prosent	Vekt	Prosent	
Mandag	-0,242	0,64 %	-0,227	0,69 %	-0,088	0,30 %	0,54 %
Fredag	-0,039	0,10 %	-0,324	0,99 %	-0,087	0,30 %	0,46 %
Onsdag	-0,131	0,34 %	-0,162	0,50 %	-0,038	0,13 %	0,32 %
Tirsdag	-0,087	0,23 %	-0,081	0,25 %	-0,111	0,38 %	0,29 %
Torsdag	-0,092	0,24 %	-0,065	0,20 %	-0,015	0,05 %	0,16 %

Tabell 5.8 - Vekting ukedag SVM

Her ser vi at alle ukedagene er negativt vektet og det er store forskjeller på hvilken ukedag som ansees å ha størst påvirkning i hver enkelt modell. Generelt sett er mandag er den dagen som har størst negativ påvirkning ifølge SVM.

5.1.3.2. Prediksjonsytelse

Etter at modellene ble generert ut fra treningsdata så ble testdata kjørt gjennom modellene. Det ga oss følgende resultater:



Figur 5.3 - Klassifiseringer SVM

De tre modellene for SVM presterer veldig jevnt, men har en større andel falske (feilaktige) predikeringer, både positive og negative, enn BPM og Logistic Regression.

Måleparameterne for modellene ga oss følgende verdier:

		Accuracy	Precision	Recall	F-Score	AUC	Log Loss
Modell 1	Trening	0,770	0,772	0,816	0,794	0,854	0,475
	Test	0,747	0,750	0,800	0,774	0,835	0,495
	Differanse	-0,023	-0,022	-0,016	-0,020	-0,019	0,020
Modell 2	Trening	0,769	0,771	0,815	0,793	0,849	0,479
	Test	0,752	0,758	0,796	0,777	0,828	0,511
	Differanse	-0,017	-0,013	-0,019	-0,016	-0,020	0,032
Modell 3	Trening	0,771	0,774	0,816	0,794	0,853	0,472
	Test	0,753	0,757	0,800	0,778	0,836	0,506
	Differanse	-0,018	-0,017	-0,016	-0,016	-0,017	0,034
Gjennomsnitt	Trening	0,770	0,773	0,816	0,794	0,852	0,475
	Test	0,751	0,755	0,799	0,776	0,833	0,504
	Differanse	-0,019	-0,017	-0,017	-0,017	-0,019	0,029

Tabell 5.9 - Ytelse SVM

SVM presterer noe dårligere enn LR og BPM når det kommer til å predikere korrekt. Den har også litt større differanse mellom trening og testresultatene mot BPM, noe som tyder på at den generaliserer marginalt dårligere.

5.1.4. Sammenligning lineære metoder

I dette delkapitlet vil vi se på hvordan de ulike variablene har blitt vektet i de ulike metodene og diskutere sammenhengen mellom de ulike metodene og variablene. Selv om modellene forklarer at de ulike variablene påvirker OSEBX i ulik grad er ikke dette nødvendigvis forklaringen på hva som påvirker Oslo Børs direkte. Modellene påpeker hovedsakelig at disse variablene korrelerer med hverandre, men det kan være utelatte variabler som påvirker Oslo Børs. Begrunnelsen på hvorfor enkelte variabler har fått tillagt større vekt enn andre variabler kan være at en eller flere utelatte variabler påvirker Oslo Børs og de uavhengige variablene på lik måte. Disse utelatte variablene kan være variabler som er vanskelig å kvantifisere. Eksempler på dette kan være nyheter og politiske vedtak. Tabell 5.10 viser en oversikt over gjennomsnittlig vektning av variablene til alle modellene til hver enkelt metode, samt et gjennomsnitt på hvilken grad av påvirkning variablene har.

Variabel	BPM	LR	SVM	Gjennomsnitt
BIAS	1,65 %	29,15 %	25,36 %	18,72 %
SWE_OMX	12,30 %	9,90 %	9,81 %	10,67 %
BRENT	11,90 %	10,37 %	7,26 %	9,84 %
DK_OMX	10,63 %	10,26 %	8,32 %	9,74 %
FTSE100	11,70 %	9,47 %	7,95 %	9,71 %
AEX	7,46 %	7,02 %	7,91 %	7,46 %
CAC40	6,10 %	6,18 %	8,32 %	6,87 %
SPCOMP_O	6,73 %	5,23 %	2,36 %	4,77 %
SPCOMP_1	5,79 %	4,15 %	3,08 %	4,34 %
DAX	2,33 %	1,67 %	7,03 %	3,68 %
OSEBX_1	3,94 %	1,77 %	1,64 %	2,45 %
EUR	3,05 %	1,52 %	2,08 %	2,22 %
GBP	3,10 %	1,14 %	0,30 %	1,51 %
SSE	1,22 %	0,59 %	1,78 %	1,20 %
VIX	0,70 %	0,23 %	2,39 %	1,11 %
USD	0,95 %	0,25 %	1,37 %	0,86 %
DATE	0,00 %	0,53 %	1,00 %	0,51 %
OSEBX_T	0,38 %	0,14 %	0,26 %	0,26 %

Tabell 5.10 - Sammenstilling vektning variabler

Det er store variasjoner mellom hvor forholdsvis stort skjevheten (BIAS) til som modellene til metodene får. Dette feilleddet er betydelig mindre ved hjelp av Bayes Point Machine enn ved lineær Regression og Support Vector Machine.

Oljeprisen er generelt sett høyt vektet i alle modellene til de ulike metodene. Det er sammenfallende med tidligere forskning som også viser at norsk økonomi i stor grad blir påvirket av svingninger i oljeprisen, se Henriksen og Killingstad (2013).

Vi ser at de ulike metodene mener at indeksene har stor innvirkning på om OSEBX går opp eller ned. Spesielt ser vi at de skandinaviske børsene og Storbritannia tillegges stor vekt.

Våre to variabler for S&P 500 har også betydelig positiv vekt. Det er sammenfallende med funnene til Svarttjernet og Ulsrud (2016). En endring i S&P 500 sin åpningspris vektlegges noe høyere enn prisendringen dagen før. Dette tyder på at modellene mener at åpningsprisen i USA er viktigere enn gårdagens endring.

Blant de resterende indeksene er de europeiske tillagt større vekt enn Kina. Det indikerer at Kina er en børs som modellene mener har mindre forklaringsgrad på bevegelsene til OSEBX enn de europeiske.

Modellene mener at valuta har noe påvirkning på om OSEBX går opp eller ned, men denne påvirkningen er forholdsmessig mye mindre enn indeksene. Euro og Pund blir lagt betydelig større vekt på enn Dollar. Dette er noe overraskende da oljekontrakter, som er en betydelig inntektskilde for Norge, omsettes i Dollar. Noe overraskende vekt alle modellene Euro negativt. Det vil si at modellene mener at dersom Euro øker vil det ha negativ innvirkning på Oslo Børs. Dette er ikke naturlig da en øking av Euro mot NOK vil øke konkurransekraften til norske eksportører. En norsk eksportør som får betalt i Euro vil få bedre lønnsomhet ved økning i Euro mot NOK, gitt at kostnadene hans er i NOK og prisen i Euro er konstant.

Volatilitetsindeksen VIX og variabelen for omsetning på Oslo Børs har gjennomsnittlig lav vekt i forhold til de andre variablene.

Ukedag	BPM	LR	SVM	Gjennomsnitt
Torsdag	3,59 %	0,18 %	0,16 %	1,31 %
Onsdag	2,57 %	0,09 %	0,32 %	0,99 %
Tirsdag	2,34 %	0,06 %	0,29 %	0,90 %
Mandag	0,75 %	0,08 %	0,54 %	0,46 %
Fredag	0,81 %	0,04 %	0,46 %	0,44 %

Tabell 5.11 - Sammenstilling vektning ukedag

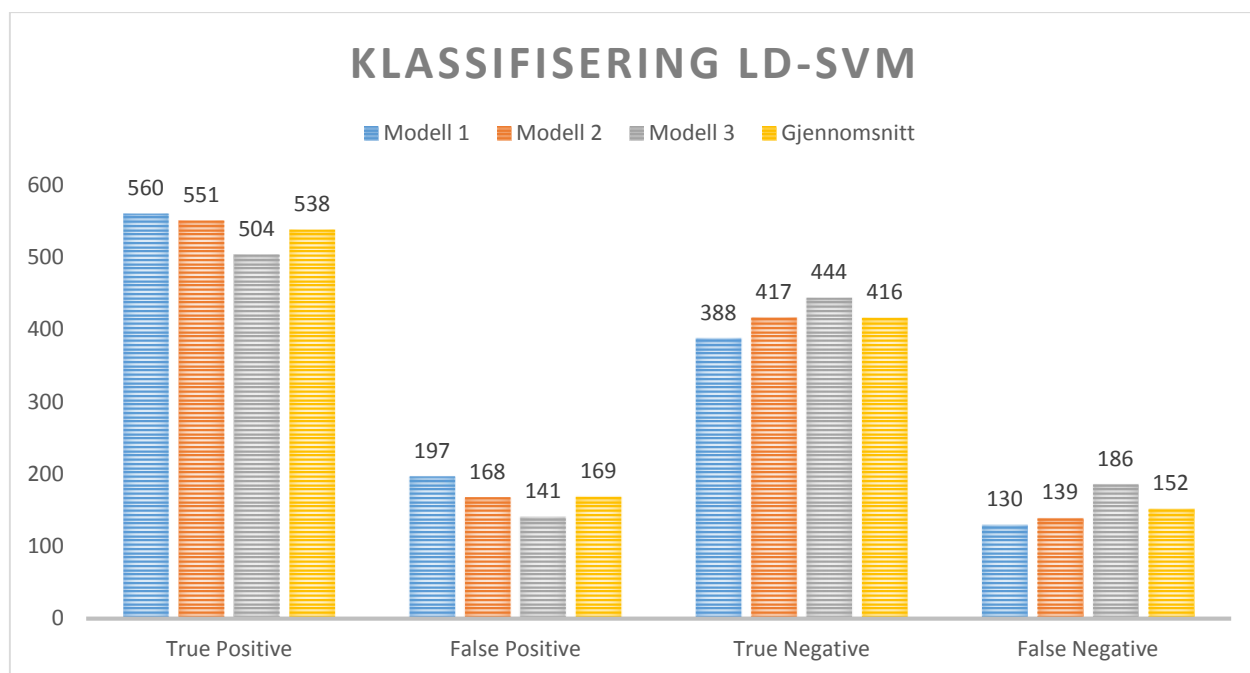
Bayes Point Machine (BPM) tillegger ukedag betydelig større vekt når det kommer til å predikere bevegelser på OSEBX. Som nevnt tidligere er feilleddet i BPM mye lavere enn i LR og SVM, men selv ekskludert feilleddet er forklaringsgraden mye større. Generelt sett virker det også som at de fleste modellene til de ulike metodene mener at det er større positiv effekt på OSEBX dersom ukedagen er andre dager enn mandag og fredag.

5.2. Ikke-lineære metoder

I dette delkapittelet vil vi se på hvordan de ikke-lineære metodene presterer. I motsetning til de lineære modellene, gir ikke de ikke-lineære metodene oss informasjon om hvilke variabler som påvirker mest prediksjonene. Dette delkapittelet vil kun analysere resultatet av prediksjonene til hver enkelte metode, og ikke inneholde en diskusjon rundt variabler slik som i det foregående delkapittelet.

5.2.1. Locally-Deep Support Vector Machine

For Locally Deep Support Vector Machine (LD-SVM) kjørte vi 30 ulike konfigurasjoner ved hjelp av hyperparameter tuning på hvert treningsdatasett. Fra hver av treningsdatasettene har vi benyttet optimal modell, og testet denne mot testdata. Det ga oss følgende resultater:



Figur 5.4 - Klassifisering LD-SVM

De tre ulike modellene har ujevn ytelse. Modell 1 har et nokså bra antall TP-klassifiseringer, men har lav andel TN. Vi ser at de andre to modellene presterer noe jevnere, men dårligere enn de tre lineære metodene. Måleparameterne for modellene ga oss følgende verdier:

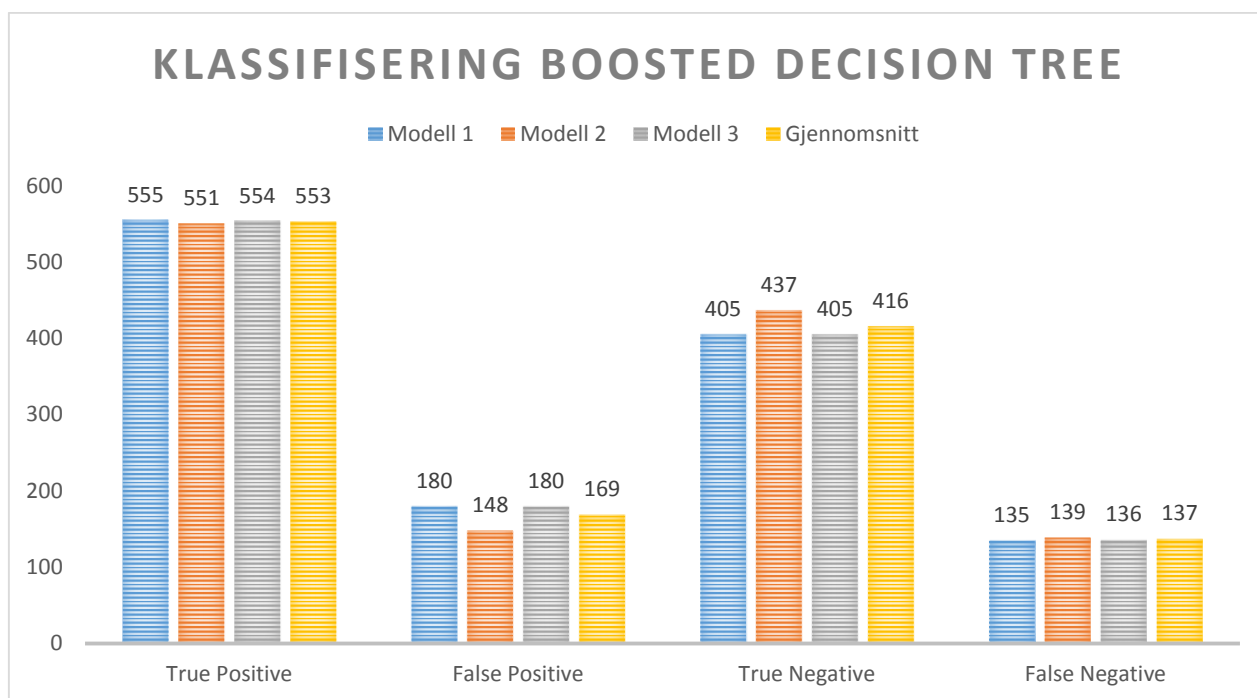
		Accuracy	Precision	Recall	F-Score	AUC	Log loss
Modell 1	Trening	0,767	0,770	0,818	0,790	0,836	0,509
	Test	0,744	0,740	0,812	0,774	0,826	0,565
	Differanse	-0,023	-0,030	-0,006	-0,016	-0,010	0,056
Modell 2	Trening	0,759	0,758	0,815	0,785	0,843	0,492
	Test	0,759	0,766	0,799	0,782	0,835	0,497
	Differanse	0,000	0,008	-0,016	-0,003	-0,008	0,005
Modell 3	Trening	0,770	0,764	0,834	0,797	0,834	0,508
	Test	0,744	0,781	0,730	0,755	0,826	0,561
	Differanse	-0,026	0,017	-0,104	-0,042	-0,008	0,053
Gjennomsnitt	Trening	0,765	0,764	0,822	0,791	0,838	0,503
	Test	0,749	0,762	0,780	0,770	0,829	0,541
	Differanse	-0,016	-0,002	-0,042	-0,020	-0,009	0,038

Tabell 5.12 - Ytelse LD-SVM

Modellene har små differanser mellom trening- og testdata - noe som tyder på at modellene generaliserer godt ut i fra treningsdata. Vi kan se ut i fra differansen på måleparameterne at modell 2 skiller seg ut i fra de andre modellene. Modell 2 har her mindre differanse mellom trening- og testdata som betyr at den har i dette tilfellet generalisert bedre ut i fra treningsdata, enn de to andre modellene. Modell 3 ser ut til å prestere noe dårligere enn 1 og 2. Den presterer spesielt dårlig på Recall, altså evnen til å identifisere faktisk positive utfall (oppgang på OSEBX).

5.2.2. Boosted Decision Tree

For Boosted Decision Tree kjørte vi 150 ulike konfigurasjoner ved hjelp av hyperparameter tuning på hvert treningsdatasett. Fra hver av treningsdatasettene har vi benyttet optimal modell, og testet denne mot testdata. Figur 5.5 viser en oversikt over klassifiseringene de tre optimale modellene leverte ut fra testdatasettet.



Figur 5.5 - Klassifisering Boosted Decision Tree

De tre modellene scorer nokså jevnt når det kommer til TP og FN, altså er modellene nokså lik i sin behandling av de dagene hvor det faktisk var oppgang. De viser noe mer ujevn prediksjonsevne på de dager med faktisk nedgang. Måleparameterne for modellene ga oss følgende verdier:

		Accuracy	Precision	Recall	F-Score	AUC	Log Loss
Modell 1	Trening	0,772	0,775	0,816	0,795	0,855	0,707
	Test	0,753	0,755	0,804	0,779	0,838	0,822
	Differanse	-0,019	-0,020	-0,012	-0,016	-0,017	0,115
Modell 2	Trening	0,768	0,779	0,799	0,789	0,860	0,466
	Test	0,775	0,788	0,799	0,793	0,849	0,482
	Differanse	0,007	0,009	0,000	0,004	-0,011	0,016
Modell 3	Trening	0,781	0,789	0,812	0,800	0,860	0,685
	Test	0,752	0,755	0,803	0,778	0,829	0,829
	Differanse	-0,029	-0,034	-0,009	-0,022	-0,031	0,144
Gjennomsnitt	Trening	0,774	0,781	0,809	0,795	0,858	0,619
	Test	0,760	0,766	0,802	0,783	0,839	0,711
	Differanse	-0,014	-0,015	-0,007	-0,011	-0,020	0,092

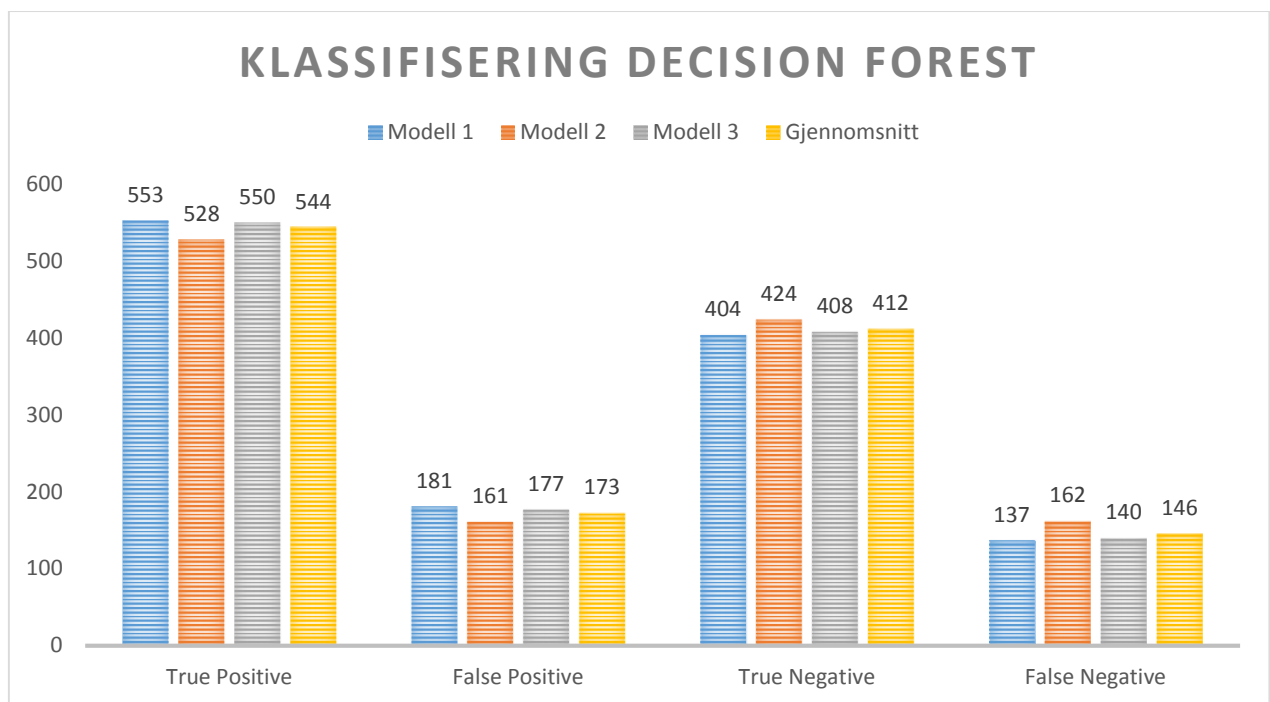
Tabell 5.13 - Ytelse Boosted Decision Tree

Modell 2 presterer markant bedre enn modell 1 og 3. Det som også er verdt å nevne er at modell 2 scorer bedre på testdata for måleparameterne Accuracy, Precision og F-score sammenlignet

med treningsdata. Et annet relevant moment når vi sammenligner modell 2 mot modell 1 og 3 er at modell 2 har betydelig lavere log loss, mens de andre måleparameterne er relativt like. Siden disse 3 modellene er generert ut i fra ulike datasett, og har fått store variasjoner log loss, tyder det på at Boosted Decision Tree som teknikk gir stor varians mellom modellene.

5.2.3. Decision Forest

For Decision Forest kjørte vi 30 ulike konfigurasjoner ved hjelp av hyperparameter tuning på hvert treningsdatasett. Fra hver av treningsdatasettene har vi benyttet optimal modell, og testet denne mot testdata. Figur 5.6 viser en oversikt over klassifiseringene de tre optimale modellene leverte ut fra testdatasettet.



Figur 5.6 - Klassifisering Decision Forest

Decision Forest spriker en del mellom de ulike modellene, men modell 1 og 3 presterer veldig jevnt. I snitt så scorer modellene nokså lavt sammenlignet med de andre metodene. Måleparameterne for modellene ga oss verdiene som fremgår av Tabell 5.14.

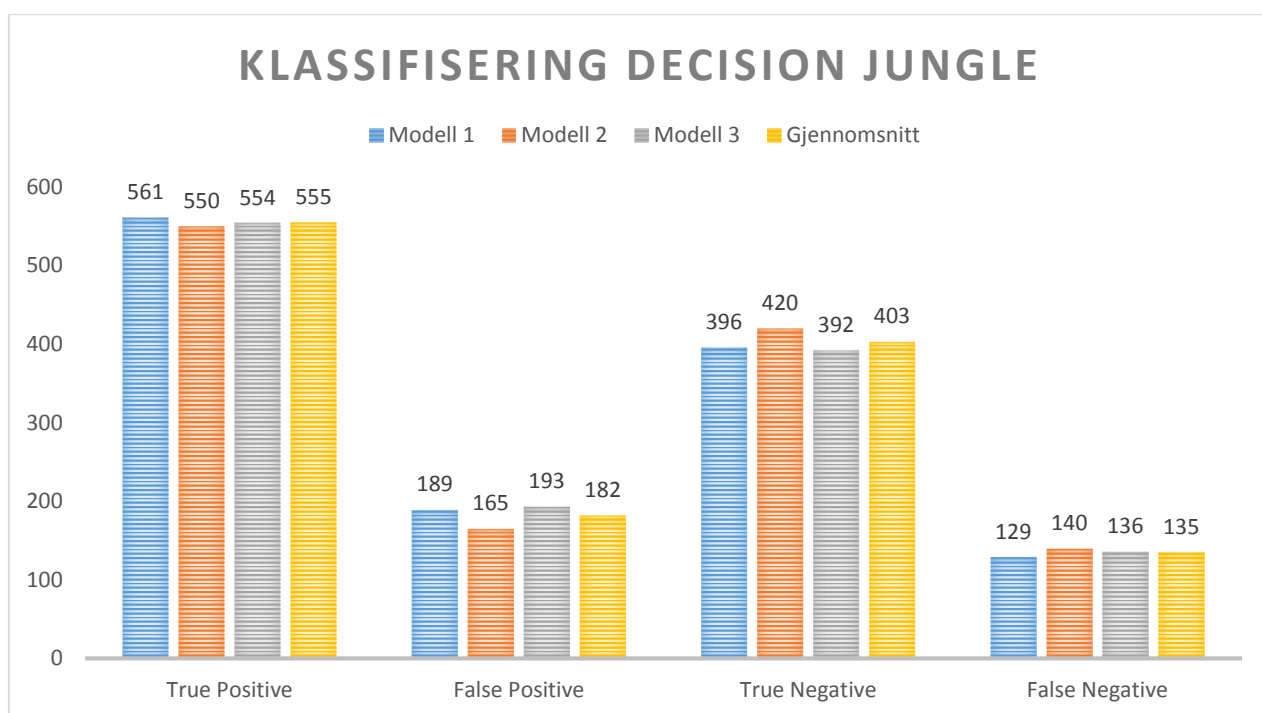
		Accuracy	Precision	Recall	F-Score	AUC	Log Loss
Modell 1	Trening	0,776	0,777	0,813	0,799	0,853	0,495
	Test	0,751	0,753	0,801	0,777	0,840	0,491
	Differanse	-0,025	-0,024	-0,012	-0,022	-0,013	-0,004
Modell 2	Trening	0,770	0,780	0,801	0,790	0,857	0,483
	Test	0,747	0,766	0,765	0,766	0,838	0,508
	Differanse	-0,023	-0,014	-0,036	-0,024	-0,019	0,025
Modell 3	Trening	0,784	0,792	0,814	0,803	0,860	0,468
	Test	0,751	0,757	0,797	0,776	0,829	0,547
	Differanse	-0,033	-0,035	-0,017	-0,027	-0,031	0,079
Gjennomsnitt	Trening	0,777	0,783	0,809	0,797	0,857	0,482
	Test	0,750	0,759	0,788	0,773	0,836	0,515
	Differanse	-0,027	-0,024	-0,022	-0,024	-0,021	0,033

Tabell 5.14 - Ytelse Decision Forest

De tre modellene presterer nokså jevnt for denne metoden, og med akseptable resultat. Modell 1 ser ut til å predikere marginalt bedre enn de to andre modellene.

5.2.4. Decision Jungle

For Decision Jungle kjørte vi 20 ulike konfigurasjoner ved hjelp av hyperparameter tuning på hvert treningsdatasett. Fra hver av treningsdatasettene har vi benyttet optimal modell, og testet denne mot testdata. Figur 5.7 viser en oversikt over klassifiseringene de tre optimale modellene leverte ut fra testdatasettet.



Figur 5.7 - Klassifisering Decision Jungle

De tre modellene scorer nokså likt på dager med faktisk oppgang, altså TP og, men i likhet med Boosted Decision Tree så scorer modellene ujevnt på dager med faktisk nedgang. Måleparameterne for modellene ga oss følgende verdier:

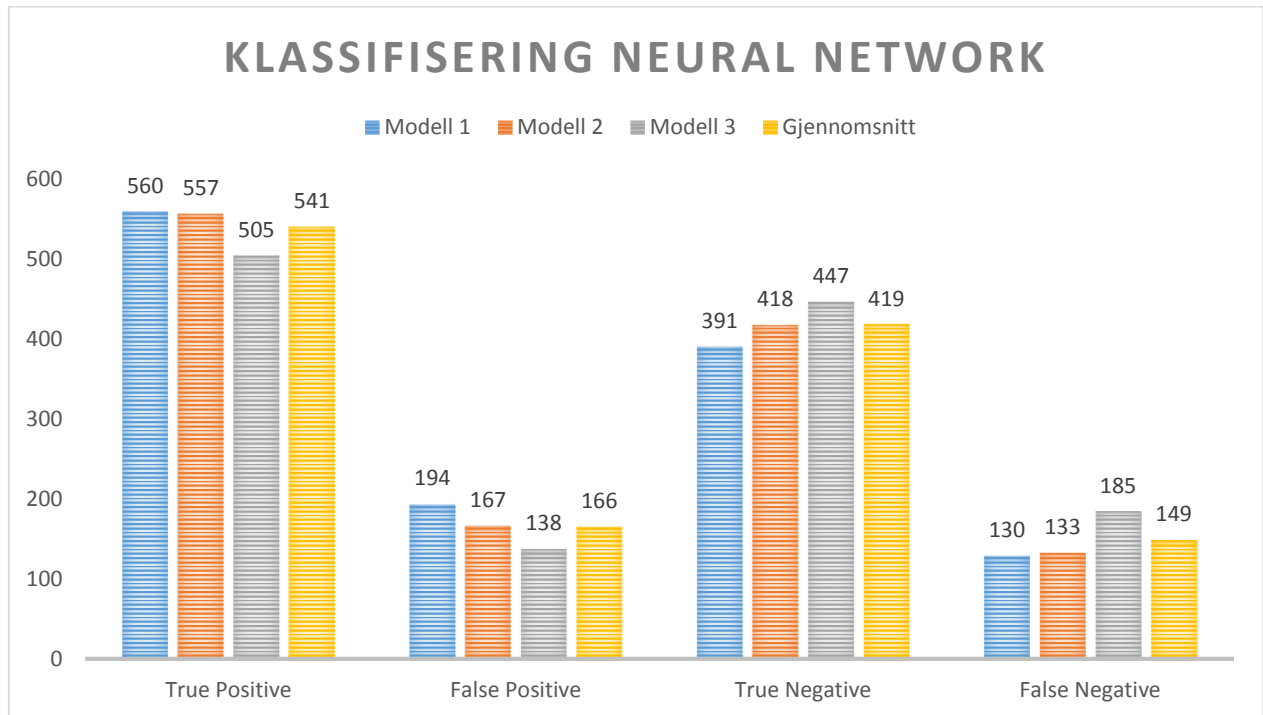
		Accuracy	Precision	Recall	F-Score	AUC	Log Loss
Modell 1	Trening	0,772	0,768	0,823	0,797	0,857	0,466
	Test	0,751	0,748	0,813	0,779	0,838	0,490
	Differanse	-0,021	-0,020	-0,010	-0,020	-0,019	0,024
Modell 2	Trening	0,768	0,777	0,801	0,789	0,856	0,475
	Test	0,761	0,769	0,797	0,783	0,843	0,482
	Differanse	-0,007	-0,008	-0,004	-0,006	-0,013	0,007
Modell 3	Trening	0,779	0,788	0,809	0,798	0,860	0,469
	Test	0,742	0,742	0,803	0,771	0,819	0,622
	Differanse	-0,037	-0,046	-0,006	-0,027	-0,041	0,153
Gjennomsnitt	Trening	0,773	0,778	0,811	0,795	0,858	0,470
	Test	0,751	0,753	0,804	0,778	0,833	0,531
	Differanse	-0,022	-0,025	-0,007	-0,018	-0,024	0,061

Tabell 5.15 - Ytelse Decision Jungle

Modell 2 predikerer bedre enn de to andre modellene. Den har bedre resultat på alle måleparameterne bortsett fra Recall, hvor modell 1 scorer marginalt bedre. Vi ser også at differansen mellom testdata og treningsdata er lavere på modell 2 sammenlignet med modell 1 og 3. Dette tyder på at Modell 2 generaliserer bedre.

5.2.5. Neural Network

For Neural Network kjørte vi 750 ulike konfigurasjoner ved hjelp av hyperparameter tuning på hvert treningsdatasett. Fra hver av treningsdatasettene har vi benyttet optimal modell, og testet denne mot testdata. Figur 5.8 viser en oversikt over klassifiseringene de tre optimale modellene leverte ut fra testdatasettet.



Figur 5.8 - Klassifisering Neural Network

Modellene generert innenfor Neural Network scorer i snitt høyest på klassiferingen TN. Men det ser ut til at modell 3 er årsaken til det. Den predikerte bare 505 av de 690 dagene med oppgang riktig. Den predikerer også nokså få FP, men er blant de høyere både på TN og FN, noe som tyder på at modellen som er generert fungerer best for å predikere nedgang. Måleparameterne for modellene en gjengitt i Tabell 5.16.

		Accuracy	Precision	Recall	F-Score	AUC	Log Loss
Modell 1	Trening	0,777	0,777	0,825	0,800	0,857	0,470
	Test	0,746	0,743	0,812	0,776	0,841	0,486
	Differanse	-0,031	-0,034	-0,013	-0,024	-0,016	0,016
Modell 2	Trening	0,771	0,802	0,768	0,784	0,858	0,471
	Test	0,765	0,769	0,807	0,788	0,840	0,490
	Differanse	-0,006	-0,033	0,039	0,004	-0,018	0,019
Modell 3	Trening	0,761	0,822	0,712	0,763	0,855	0,483
	Test	0,747	0,785	0,732	0,758	0,837	0,512
	Differanse	-0,014	-0,037	0,020	-0,005	-0,018	0,029
Gjennomsnitt	Trening	0,770	0,800	0,768	0,782	0,857	0,475
	Test	0,753	0,766	0,784	0,774	0,839	0,496
	Differanse	-0,017	-0,035	0,015	-0,008	-0,017	0,021

Tabell 5.16 - Ytelse Neural Network

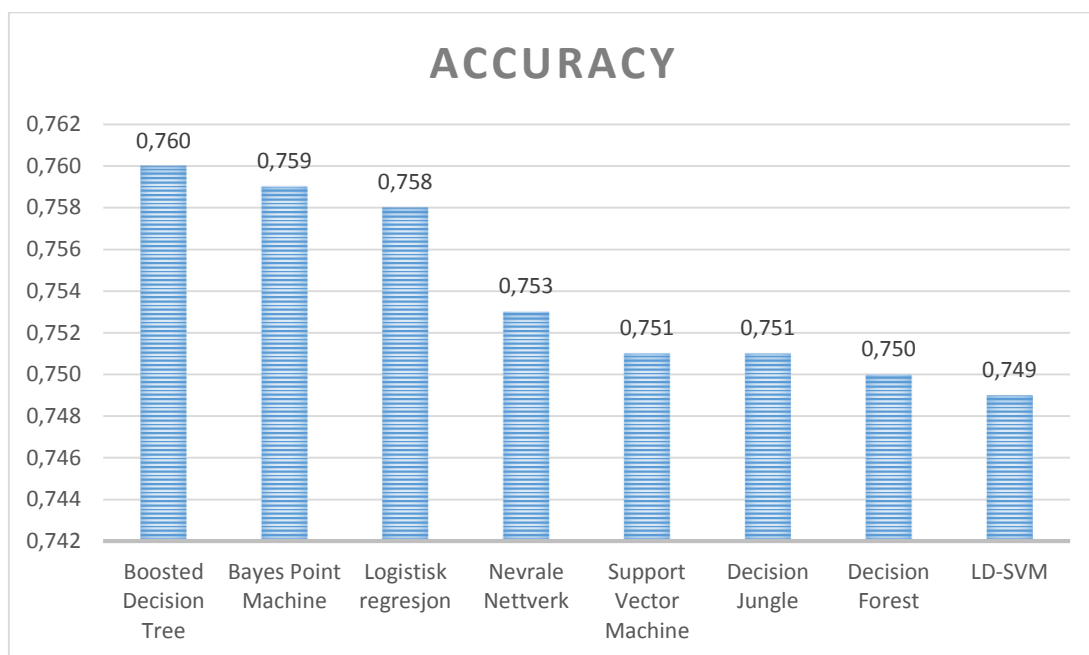
Modell 1 og 2 scorer noe høyere enn modell 3 på alle måleparametere enn Precision. Det er vanskelig å skille modellene fra hverandre, da de presterer best på ulike måleparametere. Det som er noe overraskende er at for både modell 2 og 3 øker Recall fra treningsdata til testdata. I modell 2 øker den med mer enn Precision synker, noe som gjør at også F-score går opp fra treningsdata til testdata.

5.3. Analyse av måleparameterne

I dette delkapittelet vil vi gå igjennom ytelsen, målt igjennom måleparameterne, til hver enkelt metode og sammenligne de med hverandre. Vi vil utelukkende her se på hvordan ytelsen har vært på hver enkelt metode på testdatasettet målt med gjennomsnittet av ytelsen til de tre modellene til metoden. I dette delkapittelet vil vi gå deskriptivt gjennom resultatene per måleparameter, før vi i delkapittel 5.4 sammenligner metodene mot hverandre.

5.3.1. Accuracy

Accuracy er som nevnt i kapittel 4.4.1 hvor god treffsikkerheten til prediksjonene til hver enkelt modell er.

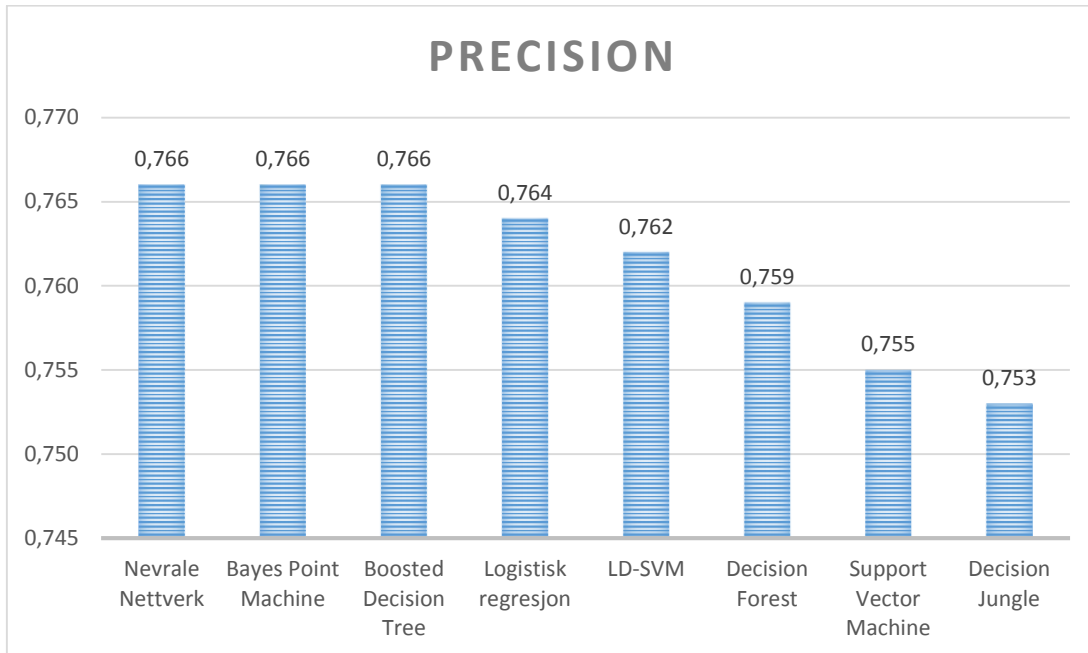


Figur 5.9 - Sammenstilling Accuracy

Som vi ser over er det marginale forskjeller i hvor god treffsikkerheten til de ulike metodene er. Dette gjør at det ikke er en klar vinner blant de ulike metodene. Den marginalt beste metoden, med hensyn på Accuracy, er Boosted Decision Tree. Den klarer i gjennomsnitt 76 % av tilfellene å fastslå hvorvidt OSEBX går opp eller ned ut fra å se på de uavhengige variablene. Dette er 1,1 prosentpoeng bedre enn modellen som scorer dårligst på denne måleparameteren, som er LD-SVM.

5.3.2. Precision

Precision er som nevnt i kapittel 4.4.2 hvor stor andel av predikeringene som er predikert positiv (oppgang på OSEBX) som faktisk var positiv.

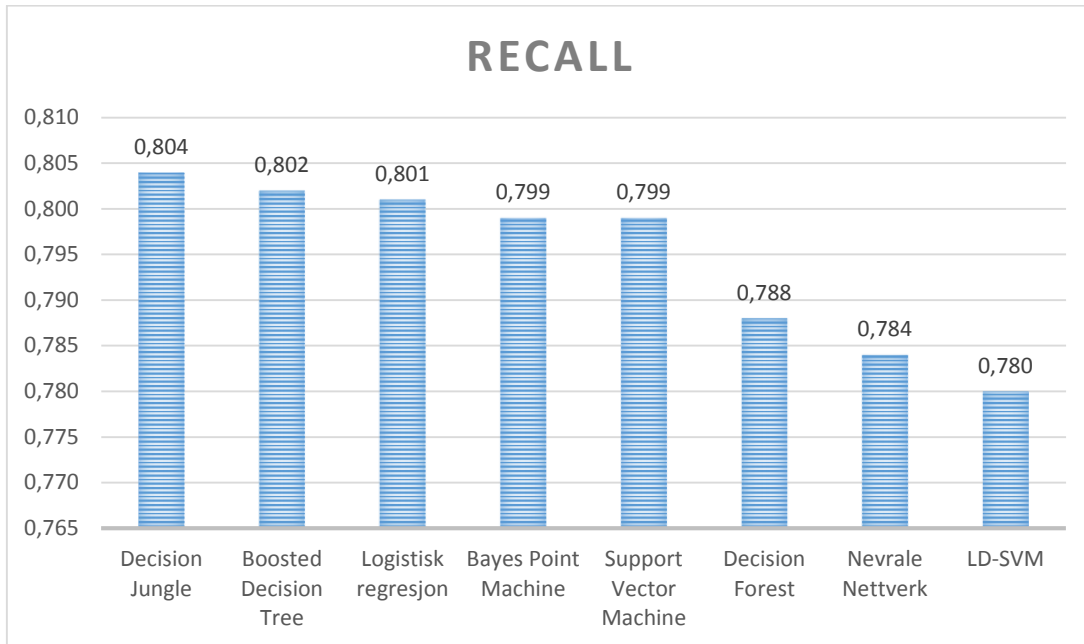


Figur 5.10 - Sammenstilling Precision

Som vi ser over er det 3 metoder som har gjennomsnittlig likt resultat. Dette gjelder Neural Network, Bayes Point Machine og Boosted Decision Tree. Andelen positive prediksjoner som faktisk var positiv for disse tre er på 76,6 %. Decision Jungle er den metoden som scorer dårligst på Precision med 75,3 %. Det er med andre ord marginale forskjeller med bare 1,3 prosentpoeng som skiller den beste og den dårligste metoden på denne måleparameteren.

5.3.3. Recall

Recall er som nevnt i kapittel 4.4.3 et mål på antall predikerte positiv (oppgang på OSEBX) sammenlignet med faktisk antall positiv.

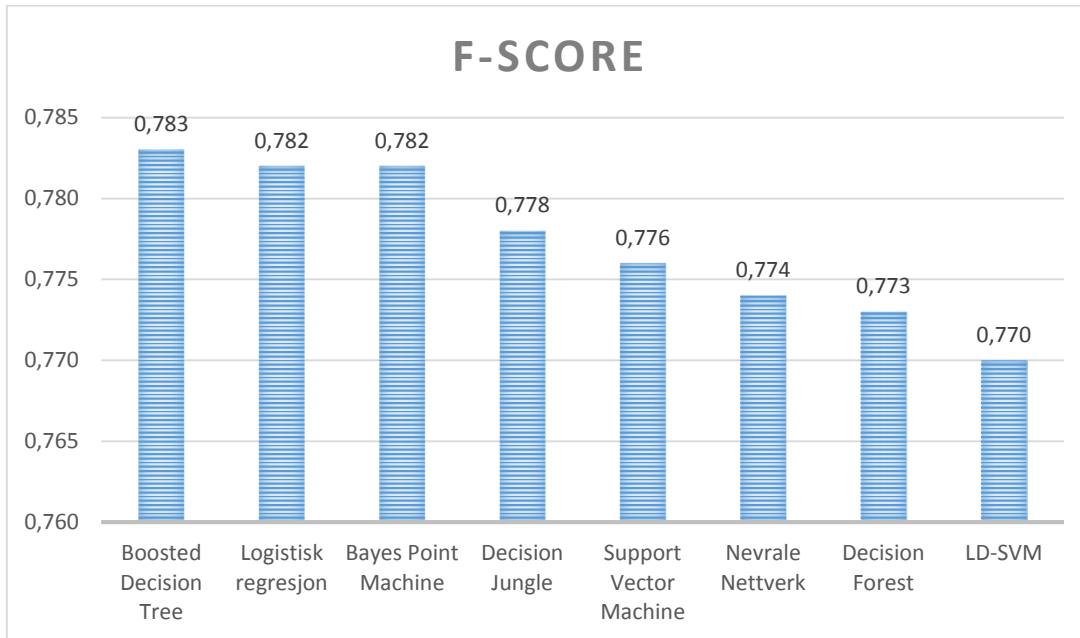


Figur 5.11 - Sammenstilling Recall

Som vi ser over er det forskjeller i hvor god Recall til de ulike metodene er. Det er ingen klar vinner i forhold til hvem som presterer best. Blant de 5 beste er det kun 0,5 prosentpoeng som skiller. Den marginalt beste metoden etter denne måleparameteren er Decision Jungle med 80,4 %, mens den dårligste, LD-SVM, har 78 %. Generelt sett ser vi også at metodene presterer bedre på Recall enn Accuracy og Precision. Siden Recall er en del høyere enn Accuracy så tyder det på at modellene fungerer bedre på å predikere oppgang enn nedgang på OSEBX.

5.3.4. F-Score

F-score er som nevnt i kapittel 4.4.4 et vektet gjennomsnitt mellom Precision og Recall. Den vil alltid være en verdi mellom 0 og 1, hvor 1 er en perfekt klassifiseringsmodell.

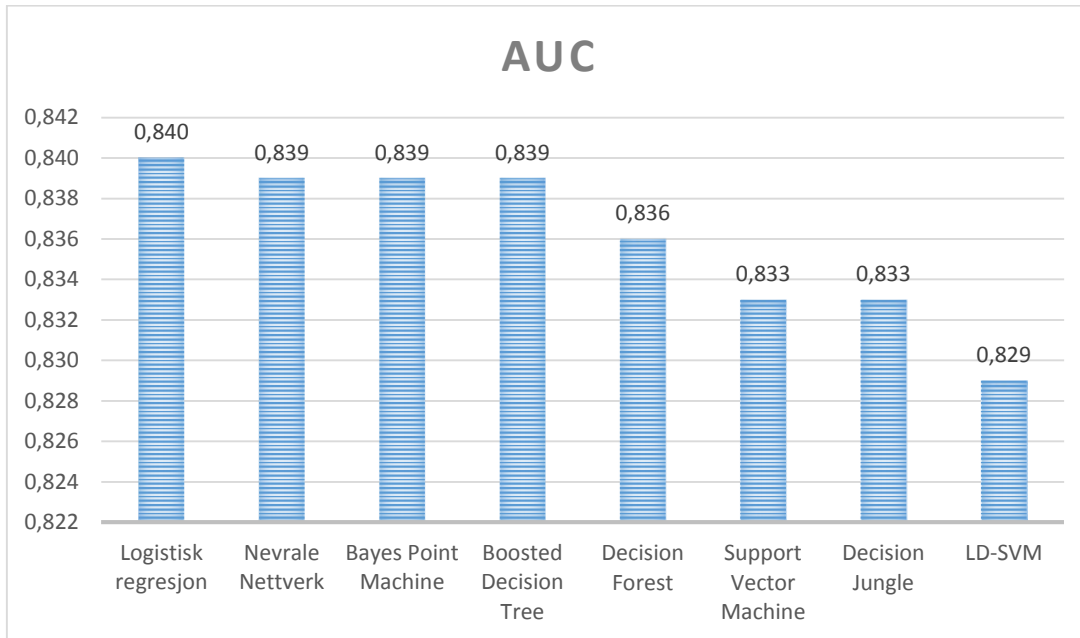


Figur 5.12 - Sammenstilling F-score

Siden dette er et vektet gjennomsnitt mellom Precision og Recall, så er resultatene på F-score også nokså høye sammenlignet med Accuracy. Det henger naturlig nok sammen med de høye resultatene på Recall. Det er lite som skiller mellom de ulike metodene, med kun 1,3 prosentpoeng som skiller den beste metoden, Boosted Decision Tree, og den dårligste, LD-SVM.

5.3.5. Area under the curve

Area Under the Curve (AUC) er som nevnt i kapittel 4.4.5 hvor stor andel av området under en ROC kurve er. Som nevnt vil en verdi på 0,5 stilsvare tilfeldig gjetting.

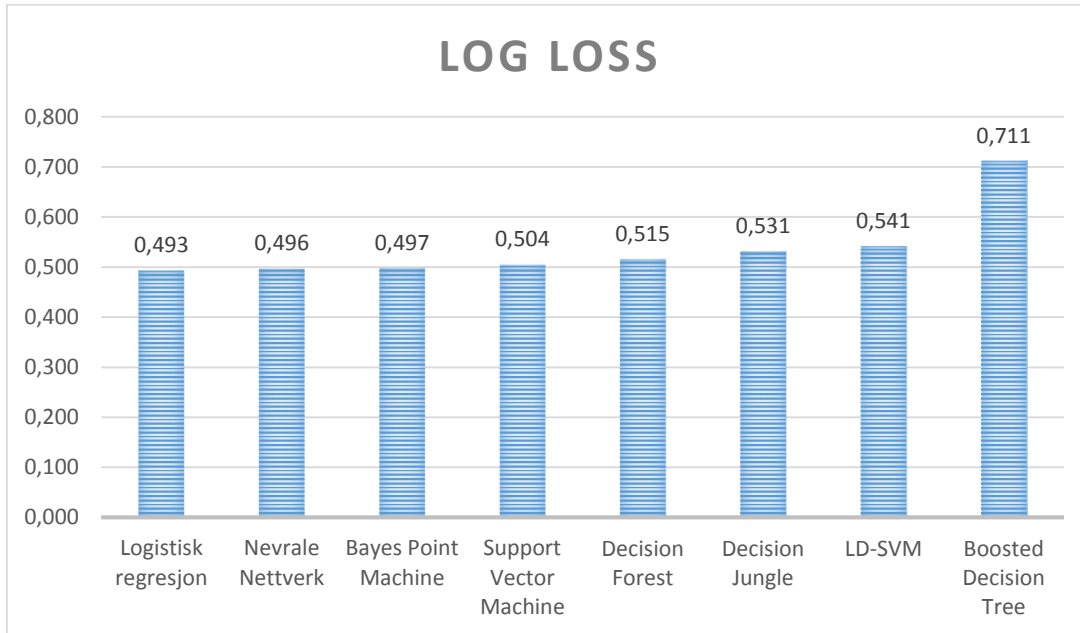


Figur 5.13 - Sammenstilling AUC

Generelt så scorer alle metodene høyt på AUC. Det er igjen lite som skiller de ulike metodene målt i AUC. Det er vanskelig å kåre en klar vinner etter denne måleparameteren da det er kun 0,1 prosentpoeng som skiller den beste fra de tre neste metodene. LD-SVM er igjen markant dårligere enn de resterende metodene.

5.3.6. Logarithmic Loss

Logarithmic Loss (log loss) tar også hensyn til sannsynligheten til prediksjonen, se kapittel 4.4.6. I motsetning til de tidligere måleparameterne er en lavere verdi å foretrekke her fremfor en høyere verdi.



Figur 5.14 - Sammenstilling Log loss

Det som kommer klart frem over er at Boosted Decision Tree, som har gode resultater på de andre måleparameterne, skiller seg markant negativt ut i forhold til de andre metodene på log loss. Den scorer høyere enn 0,693, som er log loss til en tilfeldig klassifikatormodell. Siden Boosted Decision Tree har scoret best på Accuracy, men dårligst på log loss, så tyder det på at den har gjort klassifiseringer med høy sikkerhet også i de tilfellene hvor den har tatt feil. Det fører til at modellen straffes i form av høy log loss.

De andre modellene scorer nokså likt, og har en akseptabel log loss-verdi, godt under 0,693. Det tyder på at disse modellene tar forbehold i form av lavere sannsynlighet for oppgang/nedgang i de tilfellene de predikerer feil.

5.4.Sammenligning av metodene

Når vi skal svare på spørsmålet hvilken metode som egner seg best er det mange faktorer å ta hensyn til. Hvilken måleparameter som skal være styrende vil variere ut fra hva som er viktig for den enkelte. Generelt sett så vil styrende måleparameter være avhengig av hvilket problem som skal løses.

Vi har valgt å legge jevn vekt på måleparameterne Accuracy, Precision, Recall, F-Score og AUC, og har derfor valgt å se på et gjennomsnitt av disse. I Tabell 5.17 - Sammenstilling av modellenes totale ytelse Tabell 5.17 har vi fremstilt en gjennomsnittlig verdi for de nevnte måleparameterne representert ved gjennomsnittlig score. I tillegg viser tabellen log loss for hver metode. Ut i fra de to verdiene så har vi generert en ratio-verdi ved å dele gjennomsnittlig score på log loss. Vi har med andre ord lagt større vekt på log loss enn de andre måleparameterne. Grunnen til dette er fordi vi mener at modeller med lavere log loss er mer troverdige, da de tar forbehold i de tilfellene hvor de er usikre. Modeller med lav log loss vil med andre ord gi høyere verdi til de som skal benytte modellene.

Metode	Gjennomsnittlig Score	Log Loss	Ratio
Logistic Regression	0,789	0,493	1,600
Bayes Point Machine	0,789	0,497	1,588
Neural Network	0,783	0,496	1,579
Support Vector Machine	0,783	0,504	1,554
Decision Forest	0,781	0,515	1,517
Decision Jungle	0,784	0,531	1,476
LD-SVM	0,778	0,541	1,438
Boosted Decision Tree	0,790	0,711	1,111

Tabell 5.17 - Sammenstilling av modellenes totale ytelse

I vårt forsøk er det små forskjeller i mellom de gjennomsnittlige måleparameterne. De to metodene med høyest ratio er de lineære metodene Logistic Regression og Bayes Point Machine. De scorer jevnt over høyt på alle måleparameterne, og ender opp med lik gjennomsnittlig score. Logistic Regression har litt høyere ratio, da den har noe lavere log loss.

Neural Network, SVM, Decision Forest og Decision Jungle, scorer nokså likt på gjennomsnittlig score. Blant disse metodene rangerer Neural Network som den med høyest ratio, og har lavest log loss blant de nevnte. Det gjør at Neural Network er den ikke-lineære metoden som fungerer best på vårt datasett.

LD-SVM scorer lavest på gjennomsnittlig score og har nest høyest log loss blant metodene. Vi anser den derfor mindre egnet for å predikere utvikling på OSEBX basert på vårt datasett.

Boosted Decision Tree scorer best på gjennomsnittlig score, men har klart høyest log loss. Dette gjør at ratioen blir dårligst av alle, og vi anser den derfor blant de dårligst egnede metodene i vår studie. Den scorer sannsynligvis dårlig på log loss fordi den gjør prediksjonene med større sannsynlighet enn de andre metodene, selv når den ikke treffer med prediksjonen. Det er uheldig hvis beslutningstakeren også baserer seg på metodens angitte sannsynlighet i sitt beslutningsgrunnlag. Dersom en prediksjon om oppgang har sannsynlighet 90 % så er det mer fristende å tro på den enn dersom den har 50 % sannsynlighet, selv om utfallet i begge tilfeller kan være nedgang.

5.5. Oppsummering

Alle metodene vi har benyttet har det vi anser som god prediksjonsevne på vårt datasett. Ved å se på dagens endringer på variablene så klarer de fleste modellene i ca. 75 % av tilfellene å predikere oppgang eller nedgang på OSEBX ved dagens slutt.

Det er vanskelig å konkludere med at 75 % er et bra eller dårlig resultat, da vi har ikke funnet noen lignende forsøk tidligere i en norsk sammenheng. Resultatene tyder på at våre valgte variabler har sammenheng med OSEBX, og at OSEBX reagerer på endringer i disse variablene. Likevel kan vi ikke utelate at andre variabler enn de vi har valgt også har sammenheng. Dette kan være utelatte variabler som påvirker både våre valgte variabler og OSEBX. Det kan være slik at våre valgte variabler ikke nødvendigvis har en direkte påvirkning, men reagerer på lik måte som OSEBX.

Ettersom vi har valgt en klassifiseringsmodell så får vi enten galt eller riktig svar, så kan det være at modellene har bommet med liten margin. Måleparameteren Accuracy sier ikke noe om dette, og derfor er det naturlig å se på log loss-resultatene. I alle metodene bortsett fra Boosted Decision Tree så har vi fått en akseptabel log loss, altså at prediksjonene er gjort med en akseptabel grad av sannsynlighet.

Alle metodene har, for vårt datasett, nokså lik prediksjonsevne på alle måleparameterne bortsett fra log loss, hvor Boosted Decision Tree skiller seg ut i negativ forstand. Vi ser at de lineære metodene jevnt scorer noe bedre enn de ikke-lineære. Dette kan tyde på at det er et lineært forhold mellom flere av våre uavhengige variabler og OSEBX. En annen fordel med de lineære metodene er at modellene som genereres er lettere å tolke og forstå da variablene blir vektet.

Neural Network er den eneste ikke-lineære modellen som presterer bedre enn en lineær modell, SVM. Neural Network presterer noe ujevnt blant de tre genererte modellene, noe som tyder på

at den har høyere varians enn de lineære metodene. De to lineære modellene som scorer høyest er Logistic Regression og BPM. De scorer likt på gjennomsnittlig score, men Logistic Regression har noe lavere log loss. Vi anser derfor Logistic Regression til å være den metoden som fungerer best til å klassifisere oppgang eller nedgang på OSEBX med vårt datasett.

6. Konklusjon

Dette studiet undersøker hvorvidt det er mulig å benytte maskinlæring for å predikere utviklingen på Oslo Børs hovedindeks ved hjelp av et utvalg variabler, herunder blant annet aksjeindekser og valutakurser. Vi har benyttet oss av daglige observasjoner i perioden 2001-2017. Vi har definert tre forskningsspørsmål som vi ønsker å svare på.

I hvor stor grad kan vi ved hjelp av maskinlæring predikere om Oslo Børs, representert ut i fra hovedindeksen, vil gå opp eller ned i løpet av en dag ut i fra dagens endringer i utvalgte variabler?

Vi har til sammen testet åtte metoder for å forsøke å predikere hvorvidt OSEBX vil gå opp eller ned ved dagens slutt. Hver av de åtte metodene har generert modeller på bakgrunn av tre ulike utvalg av vårt datasett. Deretter ble modellene testet mot resterende data fra samme datasett for å forsøke å klassifisere oppgang eller nedgang. Det vil si at det totalt ble generert 24 ulike modeller. De modellene som har best resultat klarte å predikere riktig utfall i ca. 77 % av tilfellene.

Hvilke(n) maskinlæringsteknikker fungerer best for å predikere daglige bevegelser på Oslo Børs ut i fra våre valgte variabler?

Vi har totalt testet åtte ulike maskinlæringsteknikker. Vi har benyttet tre metoder som genererer lineære modeller og fem metoder som genererer ikke-lineære modeller. Alle de åtte metodene presterte nokså likt med hensyn på å predikere korrekt utfall. Det skiller kun 1,1 prosentpoeng med hensyn på gjennomsnittlig Accuracy mellom dårligste og beste metode. Metodene presterer noe mer ujevnt på måleparameteren Log Loss, som er et kvalitetsmål på prediksjonene som er gjort. De lineære metodene vi har brukt presterer noe bedre enn de ikke-lineære. I vårt forsøk er det Logistic Regression som prester best, tett fulgt av Bayes Point Machine og Neural Network.

Er det mulig, ved hjelp av maskinlæring, å forstå hvilke av våre utvalgte variabler som har størst påvirkning på Oslo Børs, representert ved Hovedindeksen?

Maskinlæringsmodeller blir som nevnt i 2.2.1 beskrevet som en svart boks som er vanskelig å analysere. Likevel gir de lineære modellene en forståelse av de uavhengige variablene siden de blir vektet i forhold til hvilke som påvirker den avhengige variabelen mest. I vårt forsøk så vi at spesielt de skandinaviske børsindeksene, FTSE100 og oljepris var høyest vektet, deretter følger S&P500 og de resterende europeiske indeksene.

6.1. Kritikk til vårt forskningsdesign

Ett problem med maskinl ring er at de genererer modeller som er vanskelige   forklare, alts  at de genererer svart boks-modeller. Man kan ha gode datasett og generere gode modeller ut fra datasettene. Modellene som genereres er i mange tilfeller multidimensjonale, som gjør at det er vanskelig   forklare eller visualisere sammenhengene. Denne problematikken kan unng s til en viss grad ved   benytte line re metoder, hvor man f r en modell som indikerer vektning av variablene man har benyttet.

I v rt fors k har vi benyttet r data er samlet inn fra akt rer som benyttes i annen forskning, ogs  p  h yere niv  enn MBA-oppgaver. Det gjør at eventuell kritikk til datasettet m  knyttes til v r behandling av r data. Vi har valgt   se bort fra differanser i noteringstid p  de europeiske b rsene, da mesteparten av  pningstiden er i samme tidsrom. Det kan tenkes at endringer som skjer etter sluttnoteringen p  OSEBX er tillagt vekt for   klassifisere oppgang eller nedgang p  OSEBX, eksempelvis har FTSE sluttnotering en time etter OSEBX.

Modellene som er generert ved hjelp av de ulike metodene har som nevnt i 4.3 brukt tilfeldige utvalg av sammensetninger p  hyperparametere. Det vil alltid v re begrensninger p  hvor mange tilfeldige utvalg vi har mulighet til   teste av hyperparametere da treningstiden  ker med hvert tilfeldige utvalg. P  grunn av denne begrensningen betyr det ikke n dvendigvis metodene som har generert modellene har den mest optimale sammensetningen, og variasjoner mellom metoder kan delvis skyldes dette.

Det er vesentlig forskjell i treningstid mellom de ulike metodene. Eksempelvis brukte maskinl ringsalgoritmen Decision Forest like lang tid p    generere 20 ulike modeller p  et datasett som Logistic Regression brukte p    generere 10.000 modeller. Da vi har begrenset metodene til total treningstid (en time per metode) er det st rre sannsynlighet for at metodene med lav treningstid per modell har f tt mer optimale hyperparametere enn andre. Vurderingene v re i forhold til hvilken metode som er best er sammenlignet med lik tidsbruk og ikke antall genererte modeller.

6.2. Videre forskning

Maskinl ring er fremdeles en relativt ny forskningsmetode, som kan anvendes p  mange ulike problemstillinger innenfor flere fagdisipliner. Det har v rt en l rerik prosess, og vi ser flere mulige retninger   ta etter   ha gjennomf rt denne studien.

En av de store problemene rundt prediksjon av b rsindekser er   finne den optimale sammensetningen av uavhengige variabler. Det er tiln rmet uendelig antall potensielle faktorer som kan p virke b rsindeksene. I v r studie er det ikke n dvendigvis de mest optimale uavhengige variablene som har blitt brukt og det ville v rt interessant om andre sammensetninger av uavhengige variabler vil f re til bedre eller d rligere prediksjonsytelse.

Det kunne ogs  v rt interessant   gj re samme type analyse, men ved bruk av en avhengig variabel som ikke er bin r. Eksempelvis kan dette gj res ved   legge inn terskelverdier for OSEBX, slik at man kan klassifisere for eksempel sterk oppgang, oppgang, nedgang og sterk nedgang. Det gj r at man vil f  et mer nyansert bilde av resultatene.

Det ville ogs  v rt interessant   benytte ett lignende forskingsdesign, men med en annen avhengig variabel. Det kunne for eksempel v rt interessant   se om man klarer   klassifisere bransjespesifikke indekser eller bestemte aksjer.

Et annet prosjekt som kunne v rt interessant   gj re et lignende fors k, men med en annen hovedindeks som avhengig variabel, for   se om de samme metodene utpeker seg som de beste, eller om dette er spesielt for v rt datasett. Da m  naturlig nok de uavhengige variablene tilpasses  konomien til landet som skal unders kes.

7. Litteraturliste

Bodie, Z., Kane, A. og Marcus A.J. (2008) *Investments*. 8. utgave. Singapore: McGraw-Hill

Brynestad, E. I. og Johnsen, J. L. (2016), *Publisering av makroøkonomiske nøkkeltall - Hvordan reagerer det norske aksjemarkedet?*, Masteroppgave, Norges Handelshøyskole. Tilgjengelig fra: <http://hdl.handle.net/11250/2403759>, [Lest 12.03.2018]

Bøhren, Ø. og Michalsen, D. (2012) *Finansiell Økonomi, Teori og Praksis*. 4. utg. Bergen: Fagbokforlaget

Clarke, K. A., (2005) *The Phantom Menace: Omitted Variable Bias in Econometric Research*, Conflict Management and Peace Science – Volume 22, Issue 4, Sept. 2005 – Side 341-352

Cowles 3rd, A. og Jones, H. E. (1937) *Some A Posteriori Probabilities in Stock Market Action*. *Econometrica*. Juli 1937, Volume 5, Utgave 3. Side 280-294.

De nasjonale forskningsetiske komiteene, (u.å). *Forskningsetiske retningslinjer for samfunnsvitenskap, humaniora, juss og teologi*. Tilgjengelig fra: <https://www.etikkom.no/forskningsetiske-retningslinjer/Samfunnsvitenskap-jus-og-humaniora/Forord/>, [Lest 27.03.2018]

Finansdepartementet. (1995) *Sparing og egenkapital*. NOU 1995:16. Oslo. Tilgjengelig fra: <https://www.regjeringen.no/contentassets/ceaff11ebb184740882c7b6157c7c5c1/no/pdfa/nou199519950016000dddpdfa.pdf>, [Lest 14.02.2018]

Gleditsch, H. J og Vollan, H. A. (2012) *Langsiktige likevektsforhold mellom aksjemarked og makroøkonomi i Norge og USA*, Masteroppgave i BE305E – Finansiering og investering, Universitetet i Nordland. Tilgjengelig fra: <http://hdl.handle.net/11250/140804>, [Lest 16.03.2018]

Henriksen, L-E. og Killingstad, M (2013) *Oljeprisens betydning på aksjemarkedet*, Masteroppgave, Universitetet for Miljø og Biovitenskap, Tilgjengelig fra: <http://hdl.handle.net/11250/187556>, [Lest 01.06.2018]

Holm, J. M. B. (2007) *Ukedagseffekter på Oslo Børs*, Masterutredning innen profilen "Finansielle emner", Norges Handelshøyskole, Tilgjengelig fra: <http://hdl.handle.net/11250/167879>, [Lest 19.05.2018]

James, G., Witten, D., Hastie, T., og Tibshirani R. (2017) *An Introduction to Statistical Learning: with Applications in R*. 7. utgave. New York: Springer

Johannessen, A., Christoffersen, L. og Tufte, P.A. (2011) *Forskningsmetode for økonomisk-administrative fag*. 3. utgave. Oslo: Abstrakt forlag AS

Jose, C., Goyal, P., Aggrwal, P. og Varma, M.. (2013) *Local Deep Kernel Learning for Efficient Non-linear SVM Prediction*. Proceedings of the 30th International Conference on Machine Learning, in PMLR 28(3):486-494

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

Liavik, T. og Andorsen, R. (2013) *Oslo Børs og makroøkonomiske variabler: langsiktige likevektsforhold gjennom to tiår*, Masteroppgave i BE305E - Finansiering og investering, Universitetet i Nordland. Tilgjengelig fra: <http://hdl.handle.net/11250/140860>, [Lest 17.03.2018]

Lintner, J. (1965), *The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets*, Review of Economics and Statistics – Volume 47, Number 1 – Feb. 1965 - Side 13–37

Malkiel, B. G. og Fama, E. F. (1970), *Efficient Capital Markets: A Review of Theory and Empirical Work*, The Journal of Finance – Volume 25, Issue 2 - Mai 1970 – Side 383-417

Mund, S. (2015) *Microsoft Azure Machine Learning*, Birmingham: Packt Publishing

Oslo Børs, (2018) *Om OSEBX*, Oslo Børs. Tilgjengelig fra: <https://www.oslobors.no/markedsaktivitet/#/details/OSEBX.OSE/overview>, [Lest 14.02.2018]

Qi, Y., Reynolds, C., Picard, R.W., (2001). *The Bayes Point Machine for computer-user frustration detection via PressureMouse*. Paper presentert på konferansen Perceptive User Interfaces, Orlando, Florida, USA, 15.-16.11.01. Konferansepaper

Sharpe, W. F. (1964), *Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk*, The Journal of Finance – Volume 19, Number 3 – Sept. 1964 – Side 425-442

Shotton, J., Sharp, T., Kohli, P., Nowozin, S., Winn, J. og Criminisi, A. (2013) *Decision Jungles: Compact and Rich Models for Classification*, Microsoft Research

Statistisk sentralbyrå. (2018). *Utenrikshandel med varer*. Tilgjengelig fra: <https://www.ssb.no/utenriksokonomi/statistikker/muh/aar>, [Lest 24.03.2018]

Svarttjernet, Ø. og Ulrud, J. (2016) *Makroøkonomiske faktorerers påvirkning på Oslo Børs*, Masteroppgave i Finansiell Økonomi, Norges Handelshøyskole. Tilgjengelig fra: <http://hdl.handle.net/11250/2407191>, [Lest 14.03.1018]

Varian, H. R. (2014), *Big Data: New Tricks for Econometrics*, Journal of Economic Perspectives - Volume 28, Number 2 - Spring 2014 - Side 3–28

Verdipapirhandelloven (2007) *Lov om verdipapirhandel av 29.06.2007 nr. 75*. Tilgjengelig fra: https://lovdata.no/dokument/NL/lov/2007-06-29-75/*#*, [Lest 14.02.2018]

DATE	OSEBX	DAY_OF_WEEK	OSEBX_1	BRENT	AEX	SSE	DAX	DK OMX	CAC40	FTSE100	SPCOMP_1	SPCOMP_0	SE OMX	USD	EUR	GBP	VIX	OSEBX_T
02.01.2001	1		2	-0,0028	0,0376	-0,0054	0,0147	-0,0223	0,005	-0,0215	-0,0077	0	-0,0137	0	0	0	27,93	1049134009
03.01.2001	0		3	0,0175	0,0004	-0,0068	0,0099	0,0231	0,0082	-0,0198	-0,0218	-0,0024	-0,0425	-0,0103	0,001	-0,0007	29,63	1292022320
04.01.2001	1		4	-0,0331	0,0482	0,0145	-0,0033	-0,0091	0,0118	0,0232	0,0241	-0,0013	0,0683	0,0023	-0,0052	-0,0053	26,35	1649860067
05.01.2001	1		5	0,0195	0,0081	-0,005	0,0038	0,0009	0,0027	-0,01	0,002	0,0007	0,0061	-0,0067	0,0024	-0,0012	27,81	1922902805
08.01.2001	1		1	0,0067	-0,0008	-0,0073	-0,0107	0,0015	-0,0019	-0,0044	-0,0078	-0,0016	-0,0056	0,0059	0,0008	0,0041	29,98	1000289060
09.01.2001	0		2	0,0023	-0,0251	0,003	-0,0004	0,0019	0,0001	-0,0121	-0,01	0,0024	-0,0221	0,0043	-0,0058	-0,0037	29,47	1345761279
10.01.2001	0		3	-0,0087	0,0352	-0,004	0,0114	-0,0132	-0,0044	-0,0017	-0,0046	-0,0025	-0,0094	-0,0066	-0,0054	-0,0044	28,44	1746291731
11.01.2001	0		4	-0,0061	0,0248	0,0107	-0,003	0,023	-0,0035	0,0087	0,0091	-0,0027	0,023	-0,0123	-0,0006	-0,0072	26,68	2193460349
12.01.2001	1		5	-0,0028	0,0012	0,0041	-0,0068	0,0038	0,0157	0,0231	0,0083	-0,0002	0,0154	-0,0041	-0,0018	-0,0108	25,31	2217416155
15.01.2001	1		1	0,0064	0,0156	0,002	-0,0344	0,0051	0,0053	-0,0002	0,0008	0	0,0129	0,0156	0,0032	0,0063	25,31	1482547039
16.01.2001	0		2	0,0011	-0,0196	-0,0145	0,007	-0,0033	-0,0021	-0,0123	-0,0141	0,0002	-0,0259	0,0013	-0,0004	-0,0033	25,59	1367862657
17.01.2001	1		3	-0,0096	-0,049	0,0166	-0,0055	0,0233	0,0263	0,0212	0,0188	0	0,0395	0,0035	0,0018	0,0083	24,4	2203519962
18.01.2001	1		4	0,0201	0,0194	-0,0128	0,0041	-0,0026	0,0061	-0,0041	0,002	0,0003	0,009	-0,0004	0,0005	-0,0025	24,48	2351481454
19.01.2001	0		5	0,0008	0,0627	-0,0082	0,011	0,0024	-0,0002	-0,0025	-0,0001	0,0049	-0,0032	0,0004	0	0,0008	23,62	3118042628
22.01.2001	0		1	-0,002	0,0533	-0,0001	0	0,0035	0,0109	0,0063	0,0036	-0,0009	-0,0064	0,0099	-0,0019	-0,0035	24,27	1568932517
23.01.2001	1		2	-0,008	-0,0242	0,0027	0	0,0071	-0,0011	-0,0073	-0,0028	0	-0,0162	-0,0047	0,0079	0,0082	23	2063041035
24.01.2001	1		3	0,0032	0,0007	0,0086	0	-0,0023	0,0171	0,0104	0,008	0,001	0,0157	0,0098	-0,0024	0,005	21,98	2675946521
25.01.2001	0		4	0,007	-0,0037	0,0003	0	0,0031	-0,0023	0,0058	-0,0014	-0,001	0,0203	0,0125	-0,0036	0,0022	22,44	2345433692
26.01.2001	1		5	-0,0064	0,0037	-0,0056	0	-0,0048	-0,0008	-0,0015	0,0062	-0,003	-0,0425	-0,0108	-0,0019	-0,0037	22,52	1699229318
29.01.2001	0		1	0,0065	-0,0033	0,0073	0	0,0083	0,0025	-0,0064	0,0036	-0,0029	0,0062	0,002	-0,0018	-0,0001	22,8	2536073384
30.01.2001	1		2	-0,0087	-0,0234	-0,0036	0	-0,0017	-0,0014	0,005	0,0028	0,0002	0,0112	-0,0015	-0,001	-0,0001	23,07	2933062997
31.01.2001	1		3	0,0099	0,0103	0,008	0	0,0083	0,0087	0,0137	-0,0058	0,0006	0,0187	-0,0081	0,0023	-0,0086	22,94	3260887596
01.02.2001	1		4	0,013	0,0218	-0,0093	0	-0,0133	-0,016	-0,0165	-0,0073	-0,0002	-0,0127	-0,011	0,001	0,0003	22,39	2381266290
02.02.2001	1		5	0,005	0,0637	-0,0048	0	-0,0099	-0,0206	-0,0124	0,0007	-0,0002	-0,0121	-0,0011	-0,001	0,0002	20,87	1490663360
05.02.2001	0		1	0,0037	0,0121	-0,0013	-0,0278	-0,0015	-0,0075	-0,0005	0,002	0	-0,0169	-0,002	-0,0001	-0,0037	22,75	1665474117
06.02.2001	1		2	-0,0051	-0,0051	0,0031	-0,0064	0,0098	0,0114	0,005	0,0039	0	0,0108	0,0075	-0,0035	-0,0002	22,04	1976272529
07.02.2001	0		3	0,0107	0,0151	-0,0113	-0,0077	-0,017	-0,0035	-0,0171	-0,0108	0	-0,0171	0,0012	0,0015	0,0018	21,63	2053302806
08.02.2001	1		4	-0,0023	0,0386	0,0033	-0,0249	0,0088	-0,0051	0,0037	-0,0031	0,0002	-0,0109	0,0094	-0,002	-0,0005	21,46	2322680829
09.02.2001	0		5	0,0064	-0,0375	-0,0069	0,014	-0,0211	-0,0032	-0,0106	-0,0067	0	-0,0311	0,0009	0,0017	-0,0025	21,67	2231563505
12.02.2001	0		1	-0,0063	-0,0071	0,0019	0,0023	0,0104	-0,0012	0,0082	0,0125	-0,0004	0,0084	-0,0046	0,002	-0,0005	23,06	1591897483

DATE	OSEBX	DAY_OF_WEEK	OSEBX_I	BRENT	AEX	SSE	DAX	DK_OMX	CAC40	FTSE100	SPCOMP_I	SPCOMP_O	SE_OMX	USD	EUR	DBP	VIX	OSEBX_T
02.01.2001	1	2	-0,2226733	1,6700994	-0,3776227	0,9348276	-1,5002717	0,3609024	-1,4729767	-0,6562700	-0,0158000	0,0126846	-0,9436373	-0,0018391	-0,0097278	0,0048604	0,8971855	-1,1463477
03.01.2001	0	3	1,1704476	-0,0023473	-0,4740898	0,6258332	1,5161645	0,6107865	-1,3570704	-1,8389829	-2,3654690	-1,0819224	-2,8975856	-1,3436240	0,1984457	-0,1075791	1,0856563	-1,0630721
04.01.2001	1	4	-2,3020606	2,1466568	0,9935889	-0,2239013	-0,6232462	0,8919061	1,5746758	2,0111251	4,1884290	-0,5802275	4,6196877	0,2977828	-1,0922299	-0,8464675	0,7220186	-0,9403856
05.01.2001	1	5	1,3077009	0,3438311	-0,3500606	0,2331529	0,0411670	0,1812982	-0,6889050	0,1573694	-0,9053176	0,3319449	0,3997022	-0,8746506	0,4898885	-0,1878931	0,8838817	-0,8467714
08.01.2001	1	1	0,4292798	-0,0562972	-0,5085424	-0,7002676	0,0810318	-0,1779101	-0,3070962	-0,6646580	-2,2144188	-0,7170534	-0,3940893	0,7667561	0,1568110	0,6634348	1,1244591	-1,1630944
09.01.2001	0	2	0,1273226	-1,1487826	0,2011802	-0,0372172	0,1076084	-0,0217326	-0,8320833	-0,8491948	-0,1752419	1,1072915	-1,5135389	0,5583235	-1,2171340	-0,5894628	1,0679179	-1,0446474
10.01.2001	0	3	-0,6275705	1,5621996	-0,2811555	0,7223940	-0,8956557	-0,3731320	-0,1230098	-0,3962409	0,3030836	-1,1275310	-0,6519020	-0,8616236	-1,1338646	-0,7019024	0,9537267	-0,9073234
11.01.2001	0	4	-0,4491412	1,0946339	0,7317495	-0,2045891	1,5095204	-0,3028521	0,5860637	0,7529199	0,7898007	-1,2187483	1,5462899	-1,6041648	-0,1346319	-1,1516605	0,7586041	-0,7540093
12.01.2001	1	5	-0,2226733	0,0336192	0,2769758	-0,4492096	0,2338469	1,1964523	1,5678578	0,6858156	0,8485425	-0,0785327	1,0306646	-0,5359477	-0,3844401	-1,7299210	0,6067188	-0,7457959
15.01.2001	1	1	0,4086918	0,6810180	0,1322751	-2,2259272	0,3202206	0,3843291	-0,0207396	0,0567130	-0,5528672	0,0126846	0,8610511	2,0303788	0,6564273	1,0168162	0,6067188	-0,9977497
16.01.2001	0	2	0,0449706	-0,9015123	-1,0046591	0,4391491	-0,2378865	-0,1935279	-0,8457193	-1,1931042	-0,0158000	0,1039018	-1,7713515	0,1675124	-0,0929972	-0,5252117	0,6377611	-1,0370699
17.01.2001	1	3	-0,6893344	-2,2232847	1,1382896	-0,3655237	1,5294528	2,0241933	1,4383155	1,5665592	0,5128755	0,0126846	2,6657394	0,4541072	0,3649844	1,3380720	0,5058315	-0,7505603
18.01.2001	1	4	1,3488769	0,8518593	-0,8875204	0,2524651	-0,1913776	0,4468001	-0,2866422	0,1573694	0,1604251	0,1495104	0,5964539	-0,0539473	0,0943589	-0,3967094	0,5147007	-0,6998309
19.01.2001	0	5	0,0243826	2,7985514	-0,5705570	0,6966444	0,1408290	-0,0451592	-0,1775539	-0,0187793	1,1506427	2,2475071	-0,2312603	0,0502690	-0,0097278	0,1333627	0,4193567	-0,4370113
22.01.2001	0	1	-0,1677720	2,3759439	-0,0124256	-0,0114676	0,2139145	0,8216262	0,4224313	0,2915780	-0,3514670	-0,3977931	-0,4483657	1,2878376	-0,4052574	-0,5573373	0,4914191	-0,9681320
23.01.2001	1	2	-0,5795318	-1,1083202	0,1805087	-0,0114676	0,4531033	-0,1154391	-0,5048186	-0,2452563	0,0093750	0,0126846	-1,1132509	-0,6141099	1,6348427	1,3220092	0,3506203	-0,7987243
24.01.2001	1	3	0,1890866	0,0111401	0,5870488	-0,0114676	-0,1714452	1,3057766	0,7019699	0,6606515	1,0751177	0,4687708	1,0510183	1,2748106	-0,5093442	0,8079999	0,2375379	-0,5885863
25.01.2001	0	4	0,4498678	-0,1866762	0,0151364	-0,0114676	0,1873380	-0,2091456	0,3883413	-0,1278238	0,2275585	-0,4434017	1,3631072	1,6265406	-0,7591524	0,3582418	0,2885358	-0,7019044
26.01.2001	1	5	-0,4697292	0,1460149	-0,3914037	-0,0114676	-0,3375485	-0,0920125	-0,1093738	0,5096669	-0,4353838	-1,3555742	-2,8975856	-1,4087592	-0,4052574	-0,5894628	0,2974050	-0,9234590
29.01.2001	0	1	0,4155545	-0,1686929	0,4974722	-0,0114676	0,5328329	0,1656805	-0,4434565	0,2915780	-0,1752419	-1,3099655	0,4064867	0,2587016	-0,3844401	-0,0112024	0,3284473	-0,6365426
30.01.2001	1	2	-0,6275705	-1,0723536	-0,2535935	-0,0114676	-0,1315804	-0,1388657	0,3337972	0,2244737	0,5548338	0,1039018	0,7457138	-0,1972447	-0,2179013	-0,0112024	0,3583809	-0,5004326
31.01.2001	1	3	0,6488851	0,4427393	0,5457057	-0,0114676	0,5328329	0,6498309	0,9269644	-0,4968973	0,5716172	0,2863363	1,2545545	-1,0570292	0,4690712	-1,3765396	0,3439684	-0,3880362
01.02.2001	1	4	0,8616277	0,9597591	-0,6463526	-0,0114676	-0,9022998	-1,2789618	-1,1320759	-0,6227178	-0,4857338	-0,0785327	-0,8757919	-1,4348133	0,1984457	0,0530488	0,2829926	-0,6896190
02.02.2001	1	5	0,3126145	2,8435096	-0,3362796	-0,0114676	-0,6763993	-1,6381701	-0,8525374	0,0483249	0,4457421	-0,0785327	-0,8350846	-0,1451365	-0,2179013	0,0369860	0,1144775	-0,9949670
05.02.2001	0	1	0,2233999	0,5236641	-0,0951117	-1,8010600	-0,1182921	-0,6152072	-0,0411936	0,1573694	-1,4843431	0,0126846	-1,1607427	-0,2623799	-0,0305452	-0,5894628	0,3229040	-0,9350322
06.02.2001	1	2	-0,3805146	-0,2496177	0,2080707	-0,4234601	0,6324949	0,8606706	0,3337972	0,3167421	0,2863002	0,0126846	0,7185757	0,9751887	-0,7383350	-0,0272652	0,2441898	-0,8284733
07.02.2001	0	3	0,7037864	0,6585389	-0,7841628	-0,5071461	-1,1481327	-0,3028521	-1,1729840	-0,9162990	-0,1416752	0,0126846	-1,1743118	0,1544853	0,3025324	0,2939906	0,1987350	-0,8020631
08.02.2001	1	4	-0,1883600	1,7150576	0,2218517	-1,6143759	0,5660535	-0,4277942	0,2451630	-0,2704204	-0,7207007	0,1039018	-0,7536701	1,2227024	-0,4260748	-0,0754535	0,1798880	-0,7097054
09.02.2001	0	5	0,4086918	-1,7062649	-0,4809803	0,8897659	-1,4205421	-0,2794255	-0,7298131	-0,5723896	-0,5360839	0,0126846	-2,1241477	0,1154042	0,3441671	-0,3967094	0,2031697	-0,7409454
12.02.2001	0	1	-0,4628665	-0,3395342	0,1253846	0,1365922	0,6723597	-0,1232480	0,5519736	1,0381130	-1,1318928	-0,1697499	0,5557466	-0,6010829	0,4066191	-0,0754535	0,3572722	-0,9602583

Appendiks C

Bayes Point Machine

Bayes point machine har ingen hyperparametre som kan stilles inn, men vi har her kjørt 4000 iterasjoner per gjennomsnittlig modell som skulle genereres.

Logistic Regression

For hver modell som ble valgt ut ved hjelp av Logistic Regression ble det generert 10 000 modeller på hvert treningsdatasett. Modell 1, 2 og 3 representerer de som hadde best score på sitt datasett. Tabellen under er oversikten over hyperparameterene til hver av disse modellene.

Hyperparameter	Modell 1	Modell 2	Modell 3
Optimization			
Tolerance	7,38214E-07	3,30623E-06	8,73766E-05
L1 Weight	0,5002977	0,658718765	0,9677804
L2 Weight	0,109155156	0,017074404	0,7728116
Memory Size	37	25	34

Hvor:

- Optimization Tolerance spesifiserer en terskelverdi som skal brukes når vi optimaliserer modellen. Hvis forbedringen mellom iterasjoner faller under den angitte terskelen, anses algoritmen for å være konvertert på en løsning, og trening stopper.
- L1 Weight er vekten på Lasso parameteren som skal anvendes for regularisering. Lasso reduserer størrelsen på koeffisientene til variablene opp mot en terskelverdi som vi angir, og i tillegg vil den prøve å gjennomføre utvelgelse av features i modellen. Dette gjøres ved at Lasso kan sette koeffisientene til 0, som er det samme som å ekskludere featuren fra modellen.
- L2 Weight er vekten på størrelsen på Ridge parameteren som skal anvendes for regularisering. Ridge inkluderer alle uavhengige variablene i modellen, og reduserer koeffisientene til variablene opp mot en terskelverdi spesifisert. Dess større denne terskelverdien er satt, dess større potensiell reduksjon av koeffisientene. Dette vil redusere kompleksiteten til modellen for å hindre overfitting, men en for stor verdi her vil kunne lede til underfitting. Ridge fungerer godt hvis det er uavhengige variabler som

korrelerer godt med den uavhengige variabelen da størrelsen på koeffisienten vil bli distribuert på alle de uavhengige variablene ut i fra korrelasjonen mellom dem.

- Memory Size angir mengden minne som skal brukes til L-BFGS-optimalisering. L-BFGS er en optimaliseringsalgoritme som er populær for parameterestimering. Denne parameteren angir antall tidligere stillinger og gradienter som skal lagres for beregning av neste trinn. Hyperparameteren begrenser mengden minne som brukes til å beregne neste trinn og retning. Når du angir mindre minne, er treningen raskere, men mindre nøyaktig.

SVM

For hver modell som ble valgt ut ved hjelp av Support Vector Machine ble det generert 1000 modeller på hvert treningsdatasett. Modell 1, 2 og 3 representerer de som hadde best score på sitt datasett. Tabellen under er oversikten over hyperparameterene til hver av disse modellene.

Hyperparameter	Modell 1	Modell 2	Modell 3
Num Iterations	47	9	65
Lambda	0,001471428	0,002264825	0,003156306

Hvor:

- Number of iterations refererer til hvor mange iterasjoner som har blitt gjort på datasettet før generering av ferdig modell.
- Lambda er verdien brukt som vekt for L1 regularisering. Større verdier på denne parameteren straffer mer komplekse modeller.

LD-SVM

For hver modell som ble valgt ut ved hjelp av Locally Deep Support Vector Machine ble det generert 30 modeller på hvert treningsdatasett. Modell 1, 2 og 3 representerer de som hadde best score på sitt datasett. Tabellen under er oversikten over hyperparameterene til hver av disse modellene.

Hyperparameter	Modell 1	Modell 2	Modell 3
Tree Depth	2	1	2
Lambda W	0.08246561	0,021697	0,080846
Lambda Theta	0.0276190583	0,048529	0,035822
Lambda Theta Prime	0.09437154	0,087841	0,098885
Sigma	0.4548512	0,489509	0,977767
Num Iterations	16545	10637	13312
Normalizer Type	MinMax	MinMax	MinMax

Hvor:

- Tree Depth spesifiser maksimal dybde av treet som kan opprettes av LD-SVM algoritmen
- Lambda W spesifiserer regulariseringsvekten (L2 regularisering) som brukes for å prøve å unngå overtilpasning.
- Lambda Theta angir hvor mye plass som skal ligge mellom en regionsgrense og nærmeste datapunkt. Modellen fungerer ved å partisjonere et plan inn i regioner. En lav Lambda Theta vil gjøre det slik at regionsgrenser er nært datapunkter i den trente modellen, noe som vil gi lavere treningsfeil – men kan lede til overtilpasning.
- Lambda Theta Prime er en verdi for å kontrollere krumningsgraden som er tillatt i beslutningsgrensene i modellen. Større verdier gir modellen større fleksibilitet til å lære buede beslutningsgrenser, mens mindre verdier kan begrense beslutningsgrensene til et trinnvis lineært mønster.
- Sigma er forkortelse for Sigmoid-skarphet. Større verdier av denne betyr at tanh i en lokal kernel Θ (theta) er mettet, mens en mindre verdi innebærer et mer lineært driftsområde for theta. (Optimaliseringsformel?)

- Num iterations er antall iterasjoner algoritmen skal kjøre på en tilfeldig del av observasjonene i datasettet for å konvergere på en løsning.
- Normalizer Type angir hvilken normaliseringsfunksjon som brukes på datasettet ved trening.

Boosted Decision Tree

For hver modell som ble valgt ut ved hjelp av Boosted Decision Tree ble det generert 150 modeller på hvert treningsdatasett. Modell 1, 2 og 3 representerer de som hadde best score på sitt datasett. Tabellen under er oversikten over hyperparameterene til hver av disse modellene.

Hyperparameter	Modell 1	Modell 2	Modell 3
Number of leaves	184	2	118
Minimum leaf instances	1	47	4
Learning rate	0,062014	0,33792	0,17902
Number of trees	429	99	392

Hvor:

- Number of leaves er antall blader (terminalnoder i beslutningstreet).
- Minimum leaf instances er minimum antall observasjoner som må treffe før et blad blir opprettet.
- Learning rate er en verdi mellom 0 og en som bestemmer læringsfrekvensen til maskinlæringsalgoritmen. Læringsfrekvensen bestemmer hvor raskt eller sakte eleven konvergerer på den optimale løsningen. Hvis trinnstørrelsen er for stor, kan du overskygge den optimale løsningen.
- Number of trees er antallet beslutningstrær i ensemblet.

Decision Forest

For hver modell som ble valgt ut ved hjelp av Decision Forest ble det generert 30 modeller på hvert treningsdatasett. Modell 1, 2 og 3 representerer de som hadde best score på sitt datasett.

Tabellen under er oversikten over hyperparameterene til hver av disse modellene.

Hyperparameter	Modell 1	Modell 2	Modell 3
Minimum number of samples per leaf node	7	14	1
Number of random splits per node	2858	1674	2979
Maximum depth of the decision trees	249	87	152
Number of decision trees	47	19	98

Hvor:

- Minimum number of samples per leaf node er en verdi på minimum antall observasjoner som er påkrevd for å lage en sluttnode (blad) i et beslutningstre. En økning i denne verdien vil terskelen for at en ny regel lages. Eksempelvis vil en verdi 1 her føre til at et blad vil lages selv om det kun et enkelt tilfelle som termineres i denne noden.
- Number of random splits per node er antallet tilfeldige delinger som er brukt for å konstruere hver node i hvert beslutningstre.
- Maximum depth of the decision trees er en begrensning på maksimum dybde til hver enkelt beslutningstre. En øking av dybden til et beslutningstre kan øke presisjonen til modellen med en risiko for at modellene blir overtilpasset og får vesentlig lengere treningstid.
- Number of decision trees spesifiserer antall beslutningstre som kan lages ved trening på treningsdatasettet. Flere beslutningstre kan potensielt øke ytelsen til modellen, men treningstiden øker.

Decision Jungle

For hver modell som ble valgt ut ved hjelp av Decision Jungle ble det generert 20 modeller på hvert treningsdatasett. Modell 1, 2 og 3 representerer de som hadde best score på sitt datasett.

Tabellen under er oversikten over hyperparameterene til hver av disse modellene.

Hyperparameter	Modell 1	Modell 2	Modell 3
Optimization Step Count	3145	16010	8874
Max Width	533	548	863
Max Depth	120	19	104
Ensemble Element Count	48	32	37

Hvor:

- Optimization Step Count indikerer hvor mange iterasjoner som skal gjøres på datasettet før den skal konvergere på den optimale løsningen per beslutningstre.
- Max Width er en begrensning på maksimum bredde på hvert beslutningstre.
- Max Depth er begrensning på maksimum dybde til hver enkelt beslutningstre. En øking av dybden til et beslutningstre kan øke presisjonen til modellen med en risiko for at modellene blir overtilpasset og får vesentlig lengere treningstid.
- Ensemble Element Count er antallet beslutningstrær som skal lages til ensemblet.

Neural Network

For hver modell som ble valgt ut ved hjelp av Neural Network ble det generert 750 modeller på hvert treningsdatasett. Modell 1, 2 og 3 representerer de som hadde best score på sitt datasett. Tabellen under er oversikten over hyperparameterene til hver av disse modellene.

Hyperparameter	Modell 1	Modell 2	Modell 3
Learning rate	0.0179259013	0,014619	0,010403
LossFunction	SquaredError	CrossEntropy	CrossEntropy
Number of iterations	154	120	153

Hvor:

- Learning rate er en verdi mellom 0 og en som bestemmer læringsfrekvensen til maskinlæringsalgoritmen. Læringsfrekvensen bestemmer hvor raskt eller sakte eleven konvergerer på den optimale løsningen. Hvis trinnstørrelsen er for stor, kan du overskygge den optimale løsningen.
- LossFunction er funksjonen brukt for å optimalisere det nevrale nettverket. Funksjonen brukes for å måle inkonsistens mellom forventet verdi Y og faktisk verdi y .
- Number of iterations er hvor mange iterasjoner som skal kjøres på treningsdatasettet.