VIBORG F.F

1. SEMESTERS PROJEKT

DATAANALYSE, ERHVERVSAKADAMIET DANIA

VEJLEDER: BJARNE TAULO SØRENSEN

GRUPPE 2:

MARTIN MUNK LAIGAARD

SEBSTIAN KROG HUSTED

RASMUS BECH POULSEN

DENNIS BORK KJELDSEN

Antal tegn inkl. mellemrum: 11.929 Dato: 6. januar 2026

Læsevejledning

Denne synopsis er udarbejdet i forbindelse med 1. semesterprøven på Dataanalyse uddannelsen hos Erhvervsakademi Dania. Projektet er lavet af gruppe 2, bestående af Dennis Bork Kjeldsen, Martin Munk Laigaard, Rasmus Bech Poulsen og Sebastian Krog Husted. Analyserne i synopsen baseres på mere dybdegående analyser, som kan findes i bilagene, sammen med den kommenterede kode. I synopsen vil der fremover blive brugt forkortelser, sådan at Viborg F.F. omtales som VFF, og machine learning omtales som ML.

Link til GitHub: https://github.com/rabech5/VFF-semesterprojekt

Indholdsfortegnelse

Figuroversigt

1 Problemstilling

I den moderne sport spiller dataanalyse en central rolle i beslutningsprocesser for at understøtte strategiske og kommercielle beslutninger for klubberne som virksomheder. Ved VFF kan evnen til at forudsige tilskuertal ved hjemmekampe bidrage til at optimere billetsalg, markedsføring og planlægning af ressourcer til kampene, blandt andet gennem mad, øl og personale på kampdagene. Den kommercielle afdeling har længe haft udfordringer med at forudsige billetsalget til klubbens hjemmebane, hvilket er en nøglefaktor for ressourceplanlægning, markedsføring og reducering af madspild – hvilket alt sammen har direkte indflydelse på klubbens finansielle resultat. For at imødekomme denne udfordring, har afdelingen et ønske om at udvikle en ML-model, der kan forudsige billetsalget en måned, 10 dage, 7 dage og 3 dage før kampstart ved at tage højde for faktorer som modstanderhold, vejrudsigt og akkumulerede point for sæsonen. Modellen skal fungere som beslutningsværktøj for at optimere ressourceplanlægning. For at benytte en sådan model til dens fulde potentiale, afhænger det dog også i høj grad af den praktiske implementering, som i sig selv også er dybt afhængig af klubbens datamodenhed, data governance og organisationsstruktur. Dette skaber altså et behov for at forstå hvordan en teknisk løsning af denne type kan tilpasses og implementeres til lige netop VFF's organisatoriske virkelighed. Dette projekt vil derfor ikke blot undersøge hvordan den optimale model bygges, men også hvordan det sikres at modellen kan implementeres i virksomheden, og integreres i klubbens daglige praksis på en måde hvor den skaber reel værdi som et beslutningsværktøj.

1.1 Problemformulering

Hvordan kan VFF styrke deres beslutningsgrundlag for billetsalg, markedsføring og ressourceplanlægning, ved at udvikle og implementere en optimal ML-model der kan forudsige tilskuertallet til hjemmekampe på forskellige tidspunkter før kampstart?

1.2 Undersøgelsesspørgsmål:

1. Hvilket niveau af datamodenhed har VFF på nuværende tidspunkt?
2. Hvilken ML-model forudsiger bedst tilskuerantal?
3. Hvordan kan modellen tilpasses VFF's datamodenhed og organisation for at skabe værdi i den kommercielle afdeling?

2 Videnskabsteori og metode

2.1 Videnskabsteori

For at kunne undersøge både de tekniske og forretningsmæssige aspekter af problemstillingen, anvender vi to forskellige videnskabsteoretiske tilgange. Projektet bygger derfor på et kombineret videnskabsteoretisk afsæt, hvor både positivismen og hermeneutikken spiller en rolle i hver deres del af projektet.

I den tekniske del anvendes en positivistisk tilgang til at forudsige tilskuertal ved hjælp af ML og kvantitative data. Formålet er at skabe et objektivt beslutningsgrundlag for VFF gennem målbare metrikker som MSE/RMSE. For at sikre høj reliabilitet og reproducerbarhed anvendes set.seed (James m.fl. 2023). Tilgangen er valgt, da den hjælper med verificeringen af virkeligheden gennem systematiske metoder. (Egholm 2017) For at sikre at projektets praktiske værdiskabelse undersøges VFF's organisatoriske forhold, herunder deres kommercielle behov, data governance og datamodenhed. På baggrund af den hermeneutiske tilgang forstås viden som en fortolkningsprocess betinget af aktørernes erfaringer, holdninger og praksis (Egholm 2017). Gennem tematisk kodning af interviews og præsentationer analyseres kvalitative data for at identifcere oplevede behov og barrierer i VFFs nuværende beslutningsprocesser (Hecker og Kalpokas 2024). Denne kvalitative indsigt sikrer, at ML-løsningen ikke blot fungerer teknisk, men også er meningsfuld og implenterbar i VFF's specifikke organisatoriske kontekst.

Kombinationen af disse to tilgange sikrer en helhedsforståelse, hvor den positivistiske del leverer objektive analyser og forudsigelser, mens den hermeneutiske tilgang har fokus på den praktiske anvendelse i organisationen. Samlet set gør denne tilgang det muligt at besvare problemformuleringen balanceret og understøtte VFF's ambition om øget datamodehed og udvikle sig til en datadrevet organisation.(Egholm 2017)

2.2 Metode

Dette projekt anvender kvantitative og kvalitative metoder, for at identificere faktorer der påvirker tilskuertallet ved VFF's hjemmekampe. Formålet med metoden er at skabe et datadrevet grundlag for analysen og samtidig supplere de informationer og indsigter fra VFF selv.

Den kvantitative data indhentes fra eksterne kilder som Superstats anvendt til kamp- og tilskuerdata, DMI til vejrdata og Date.nager til information om helligdage. Kilderne bidrager med vigtige variabler, der kan påvirke tilskuertallet, såsom modstanderhold, vejrforhold og kalender relaterede faktorer. For at sikre høj reliabilitet, importeres og lagres data i gruppens SQL database. I databasen renses den rå data, der håndteres manglende værdier og ensretning af formater. Den endelige databehandling og modellering foregår i Rstudio. Den kvalitative del af metoden består primært af et interview med VFF samt materiale fra klubbens præsentationer. Disse kvalitative data fungerer som et supplement til de kvantitative fund og bruges til at perspektivere resultaterne, fx ved at identificere faktorer, som ikke umiddelbart kan måles direkte i de indsamlede data.(Egholm 2017)

3 Analyse

3.1 Data Governance

Dette afsnit er baseret på den fulde data governance analyse, som kan findes i Bilag 1: Data Governance.

Analysen for VFF viser, at der er en tydelig strategisk forståelse af data som en værdiskabende ressource, der skal understøtte sportslige og kommercielle beslutninger, herunder kampdagsprædiktioner og ressourceplanlægning. Missionen om at anvende data effektivt, sikkert og struktureret er tydelig hos Daniel og Olga. Data governance er dog endnu ikke formaliseret i organisationen, men praktiseres i høj grad gennem uformelle arbejdsgange og individuel viden.

Kvalitetssikring, validering og dokumentation af data udføres primært manuelt, og der findes kun få automatiske kontroller. Dette øger sårbarheden for VFF´s datakvalitet og risikoen for at de tager beslutninger på fejlagtige grundlag. Manglen på fælles datadefinitioner, faste regler for datastrukturer og ensartet dokumentation har tidligere medført, at automatiserede løsninger er blevet vanskelige at genanvende. Organisationen mangler at implementere Data Governance og dermed fordele ansvarsområder. Daniel skal fremover agere som Data Governance Officer i VFF. Han skal skabe en håndbog med retningslinjer og datatilgang beskrivelser som skal introduceres for organisationens forskellige afdelinger.

Samlet set har VFF en klar strategisk ambition om at anvende data som beslutningsgrundlag, men manglende formaliseret data governance, personafhængighed, manuelle kontroller og et fragmenteret systemlandskab. For at kunne højne organisationens datamodenhed skal de ovenstående tiltag implementeres.

3.2 Datamodenhed

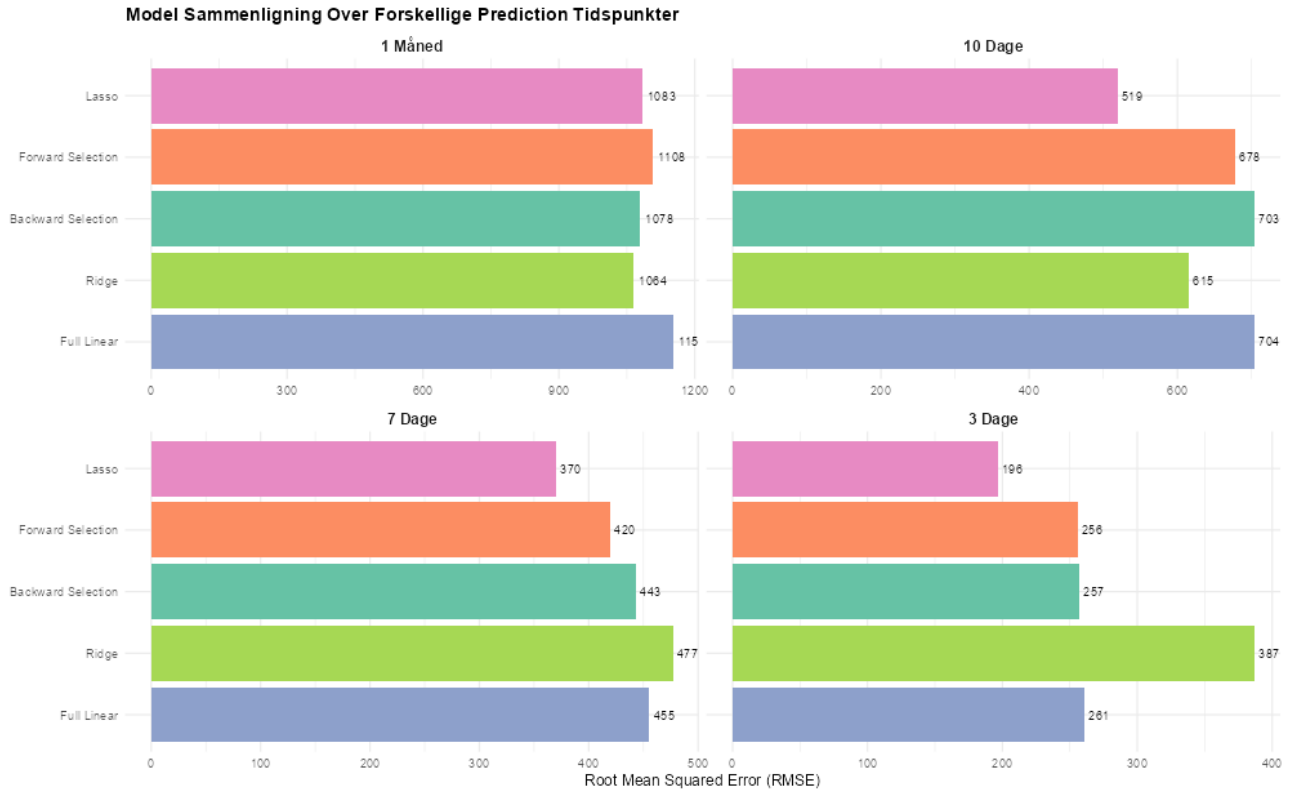
Dette afsnit er baseret på den fulde datamodenhedsanalyse, som kan findes i Bilag 2: Datamodenhed .

På baggrund af data governance-analysen og gennemførte interviews vurderes VFF at befinde sig i den accelererende del af fase 2 (Lære om forretningen) af Alexandra Instituttets datamodenhedsmodel. Organisationen anvender i stigende grad data til at træffe sportslige og kommercielle beslutninger. Dog udtrykker Daniel, at de er længere, men en lang række faktorer holder dem tilbage fra at opnå ønsket resultater. Deres tankegang og mål ligger i fase 3 og 4, men grundet silotænkning og mangelfuld datahåndteringsprocedure vil Daniel skulle påtage sig rollen som Data Governance Officer og indsætte regler for datahåndtering. Datakvalitetssikring og vedligeholdelse varetages primært af to nøglepersoner, hvilket medfører sårbarhed og begrænser skalerbarhed for VFF. Dette er ikke foreneligt med overgang til fase 3, hvor Data Governance, roller og processer er formaliseret gennem en håndbog. VFF viser en tydelig ambition om at anvende data mere som et konkurrenceparameter, men deres nuværende bemanding i dataafdelingen begrænser mulighederne for at automatisere og videreudvikle processer. Den manglende kapacitet betyder, at VFF vil forblive i den udførende del af data frem for at benytte det strategisk.

Samlet set anvender VFF data aktivt til læring og beslutningsstøtte, men manglende formaliseret Data Governance,

automatisering, systemintegration og organisatorisk kapacitet betyder, at klubben fortsat befinder sig i fase 2 og endnu ikke kan operere på fase 3-niveau.

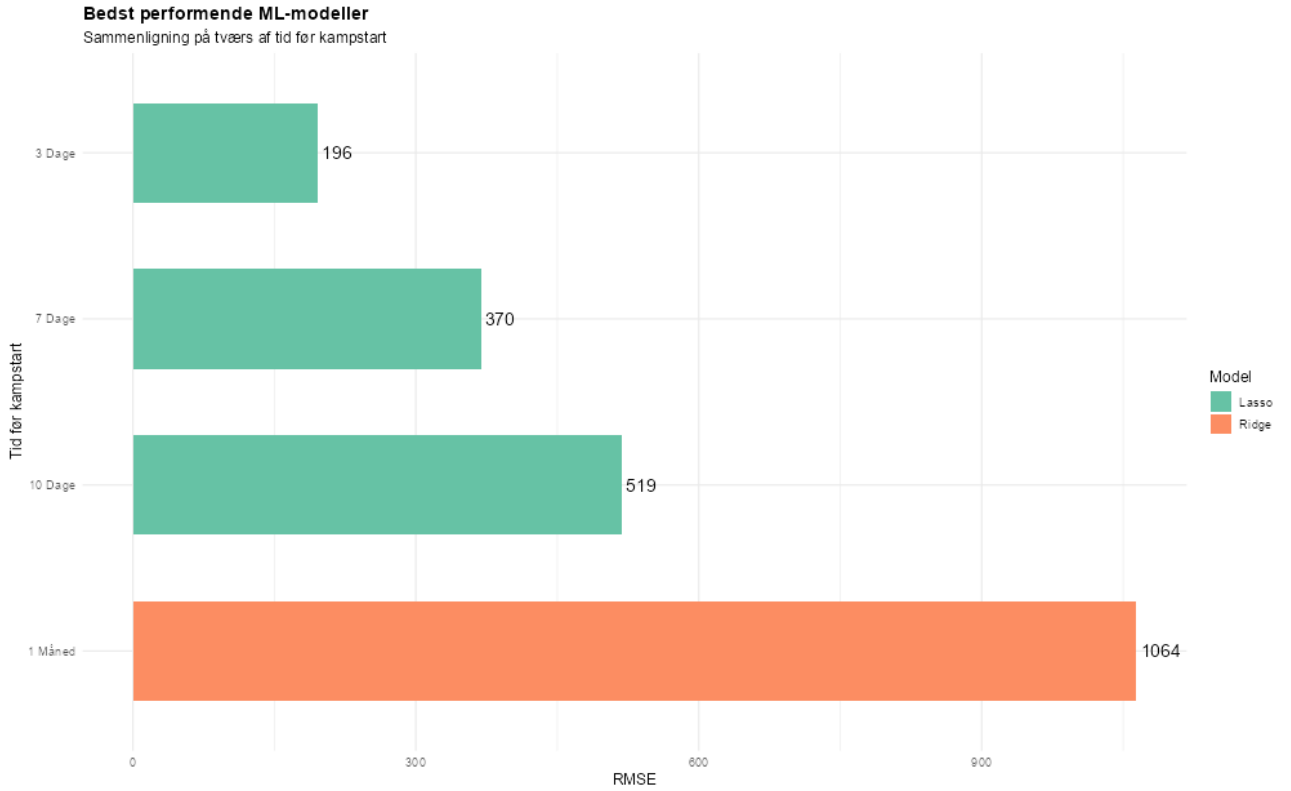
3.3 Sammenligning og vurdering af modeller



Figur 1: Model sammenligning over forskellige prediction tidspunkter (Egen tilvirkning)

Der er blevet testet og trænet 5 forskellige modeltyper på tværs af 4 forskellige tidshorisonter før kampstart. Disse modeller evalueres her på baggrund af RMSE, for at kunne sammenligne dem på den mest fortolkelige metrik (James m.fl. 2023). På tværs af modellerne, ser man et tydeligt mønster, en undtagelse ses dog ved de modeller der forsøger at forudsige 1 måned før kampstart, hvor Ridge tager føringen, da den opnår en lavere RMSE end både Lasso og de andre modeller. Dvs. at når der skal forudsiges 10, 7 eller 3 dage før kampstart, er det Lasso der performer bedst i test, mens det ved forudsigelse 1 måned inden kampstart er Ridge der performer bedst i test.

For at vurdere de bedst performende modeller fokuseres der fortsat på RMSE som evalueringsmetrik. Figuren her viser at allerede fra 1 måned til 10 dage sker der en betydelig forbedring. Denne forbedring kan primært tilskrives inkluderingen af de nye variabler for det faktiske billetsalg før kampstart. Med disse variabler får modellerne et meget præcist og retvisende pejlemærke, der skaber større præcision i forudsigelserne.



Figur 2: Bedst performende ML-modeller (Egen tilvirkning)

RMSE-værdien på 1064 ved prædiktioner en måned før kampstart indikerer en gennemsnitlig afvigelse på over 1.000 tilskuere. Uden adgang til præcise billetsalgsdata har disse langsigtede forudsigelser begrænset operational værdi, da fejlmarginerne er for store til at danne et solidt beslutningsgrundlag for eksempelvis bemanding og indkøb. Når vi derimod kigger på 10, 7 og 3 dage før kampstart ses det at RMSE'erne ligger på hhv. 519, 370 og 196 tilskuere. Denne udvikling understreger, at modellens anvendelighed er tæt forbundet med tidshorisonten. De langsigtede forudsigelser giver primært et vejledende billede, hvor de kortsigtede giver et reelt billede af forventet tilskuere. Det vurderes at de kortsigtede modeller, der inddrager billetsalgsdata tæt på kampstart, er bedst til at understøtte konkrete beslutninger om ressourceplanlægning i VFF's kommercielle afdeling.

Bilag 6: Modellering

4 Væsentligste konklusioner

Analysen placerer VFF i fase 2 af datamodenhedsmodellen. Selvom data anvendes i beslutninger, sænkes udviklingen af manglende data governance, høj personafhængighed og uklare strukturer, som dataudviklingen styres af få personer. Disse organisatoriske faktorer er afgørende for, hvordan en ML-løsning kan implementeres og skabe reel værdi.

Resultaterne viser, at modellernes præcision er tæt forbundet med tidshorisonten, da inkluderingen af billetsalgsdata markant forbedrer performance. Ridge-modellen performer bedst ved en måned før, mens Lasso-modellen leverer de mest præcise resultater tæt på kampstart, altså 10, 7 og 3 dage før kampstart. De lave RMSE-værdier tæt på kampdagen gør disse modeller velegnet som beslutningsgrundlag, mens de langsigtede fungerer som en vejledende mulighed.

Det konkluderes at VFF kan styrke deres kommercielle fundament ved at bruge ML som et værktøj tæt på kampstart. For at få fuldt udbytte af modellen skal den indføres gradvist og støttes af faste aftaler om dokumentation, ansvar og data governance. Ved at koble de tekniske forudsigelser med klubbens faktiske arbejdsgange kan VFF begrænse ressourcespild, tage mere præcise beslutninger og derfor være et skridt tættere på at blive en mere datadrevet organisation.

5 Literaturliste

Bang, Claus Grand. 2024. *Data-Driven Decision-Making for Business*. Routledge.

Data Governance Institute. u.å. "DGI Data Governance Framework Components". The Data Governance Institute. Set 2. januar 2026. https://datagovernance.com/the-dgi-data-governance-framework/dgi-data-governance-framework-components/.

DMI. u.å. "DMI Open Data API Portal". Govcloud.dk. Set 2. januar 2026. https://dmiapi.govcloud.dk/#.

Egholm, Liv. 2017. *Videnskabsteori: Perspektiver på organisationer og samfund*. Hans Reitzel.

Hecker, Jörg, og Neringa Kalpokas. 2024. "How to Code Research Interviews? | Guide & Examples". ATLAS.ti. https://atlasti.com/guides/interview-analysis-guide/coding-interviews.

James, Gareth, Daniela Witten, Trevor Hastie, og Robert Tibshirani. 2023. *An Introduction to Statistical Learning : with Applications in R*. 2nd edition. Springer.

Kølsen, Camilla, Laura Lynggaard Nielsen, og Rasmus Bækby. 2017. "Find vej i din dataindsats". Alexandra Instituttet; Alexandra Instituttet for Industriens Fond. https://alexandra.dk/wp-content/uploads/2020/09/Alexandra-Instituttet-BDBA-Find-vej-i-din-dataindsats.pdf.

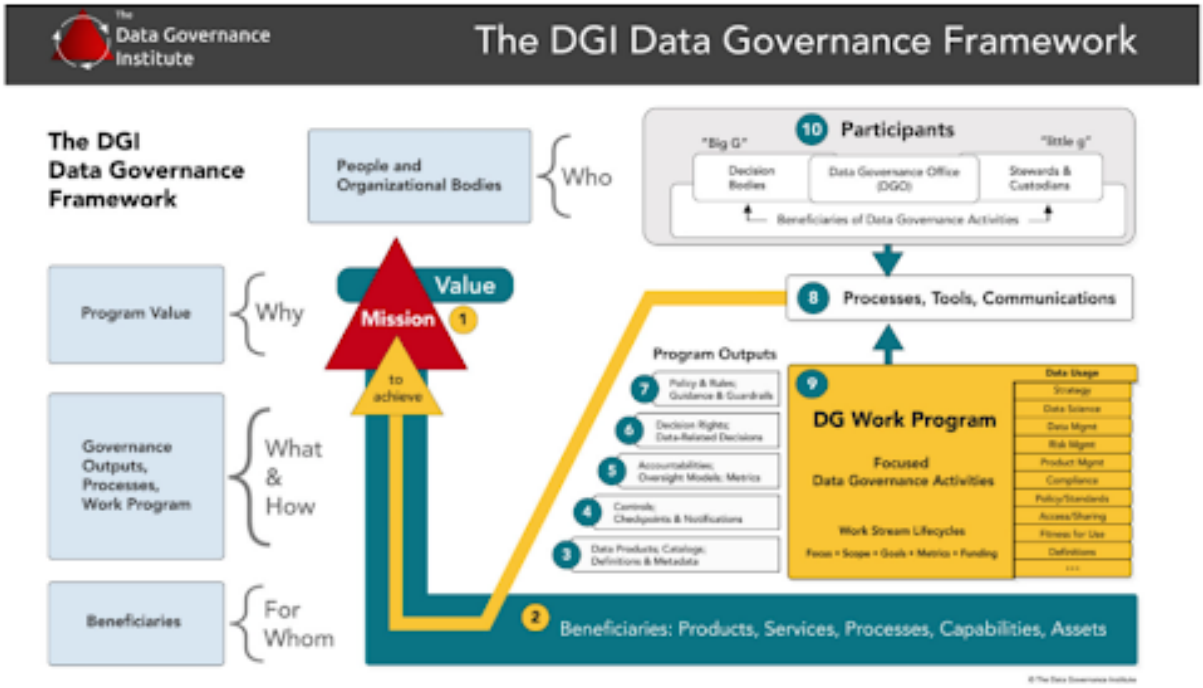
Superstats. 2026. "Superligaen i tal, fodbold, statistik, resultater og tabeller - SuperStats". Superstats.dk. https://superstats.dk.

Viborg F.F. 2019. "Viborg F.F. Prof. Fodbold A/S". Www.vff.dk. https://www.vff.dk.

Wickham, Hadley, Mine Çetinkaya-Rundel, og Garrett Grolemund. 2023. "R for Data Science". O'Reilly Media. https://r4ds.hadley.nz.

"Worldwide Public Holidays - Nager.Date". 2025. Nager.at. https://date.nager.at.

A Bilag 1: Data Governance



Figur 3: Data Governance Framework (Data Governance Institute u.å.)

A.1 Mission & Values:

Viborg F.F's mission er at sikre en effektiv, sikker og struktureret tilgang til arbejdet data, som kan styrke beslutningsgrundlaget internt i form af sportslig og kommerciel drift samt kampdags prædiktioner således at ressourceplanlægning kan gøres klar. Mission er at sikre at Viborg FF bruger data effektivt, sikkert og struktureret og dermed kan de træffe bedre beslutninger internt i form af sport, kommerciel drift og kampdags prædiktioner således at resource planlægningen kan gøres klar. (Data Governance Institute u.å.) "Vores opgave er lidt det samme. Indsamle opgaver og præsentere sådan så at Dem der skal bruge det [...] i stedet for de skal sidde i Excel og downloade og smide over og alle mulige forskellige ting, så går de bare ind og kigger, og så får de egentlig bare data præsenteret sådan så at deres opgave bliver at kigge på det færdige produkt og ikke lave det, fordi det er ikke deres kompetencer." (Bilag 4, D.2: Præsentation af Daniel)

Data anses i VFF for at være værdiskabende og målet er, at data integreres i samtlige af klubbens processer på tværs af alle afdelinger i organisationen, så data bliver en strategisk ressource, som kan bruges til at skabe en fælles arbejdsstandard. Man vil derudover gerne skabe en balance mellem automatisering og manuelt arbejde, sådan at data produkterne kommer i hænderne på dem der skal bruge det hurtigst muligt, men hvor de stadig kun skal lave så lidt manuelt arbejde som muligt: "men hvad nu hvis vi introducerer bare et manuelt step, sådan så I lidt hurtigere kan komme i gang øh med at få de ting, I godt kunne tænke jer at vide." (Bilag 4, D.2: Præsentation af Daniel) Dette er støttet af Viborg FF interviewet hvor Daniel snakker om at samle organisationen data ét sted og skabe struktur. Palle ønsker pålidelig fremmøde-data for at planlægge indkøb.

A.2 Beneficiaries of Data Governance

Hos VFF ønsker man at anvende data til at optimere sportslige resultater, både under kampe og på træningsbanen, hvor man kan monitorere spillernes belastning, for at minimere skadesrisikoen. I den kommercielle afdeling bruger man data til at udregne brugsværdier ved kampe og data fra tidligere kampe til at forudsige fremmøde til kommende kampe. Derudover bruges data til at skabe overblik over økonomi og KPI, som ledelsen anvender til at styrke beslutningsgrundlaget.

Ligeledes vil øget fitness-for-use resulterer i færre data kvalitetsfejl, minimere manuel datahåndtering og dermed give medarbejderne bedre mulighed for at finde data og have en klar opgavefordeling. (Data Governance Institute u.å.)

A.3 Data Products

Viborg FF er opbygget som et fagbureaukrati med flere uafhængige afdelinger med specialister, som ofte er autodidakte, og arbejder i forskellige systemer og egne metoder til datahåndtering. Det har skabt et stort og fragmenteret systemlandskab og forskellige tilgange til arbejdet med data. Dette resulterer i organisatoriske siloer, idet "afdelingerne nogle gange godt kan have lidt svært ved at have tid til at snakke sammen med hinanden" (Bilag 4, D.2: Præsentation af Daniel), og kan medføre at data ikke er konsistent på tværs af klubben og flere opgaver er medarbejder afhængig. Dette er også resultatet af at virksomheden er "meget sådan autodidakt oplært" (Bilag 4, D.2: Præsentation af Daniel), hvilket har skabt udfordringer med at få implementeret standarder. Dataafdelingen arbejder derfor på at standardisere processer og værktøjer med henblik på at skabe en ensartet datahåndtering, blandt andet gennem metadata, som kan bruges på tværs af organisationen. Et eksempel på denne standardisering er den sportslige sektor, hvor: "vi lavede unikke ID'er til hver eneste spiller på tværs af hele Viborg FF".[lydfil 1]

Derudover benytter klubben eksterne dataleverandører, herunder tracking data, der registrerer spillernes positioner og bevægelser, samt kampstatistikker som expected goals (xG), afslutninger, løbedistancer m.m. Disse data bruges af den sportslige afdeling til at analysere spillernes præstationer og monitorere deres belastning. VFF anvender ligeledes Eventii til billet og sæsonkort salg, sæsonkort på abonnement. Det vil sige at det er derfra kundedata fra tilskuer hentes. Det benyttes primært af Palles afdeling. Et data produkt Viborg FF bruger på nuværende tidspunkt er BI til dashboards, såsom nedenstående eksempel fra VFF powerpointen:

(Data Governance Institute u.å.)

A.4 Controls

De seneste år har medarbejderne i Viborg FF udvist en øget interesse i hvordan data kan anvendes til at optimere sportslige og kommercielle resultater. For at understøtte udviklingen har dataafdelingen udviklet et standardiseret dataprodukt template, som alle medarbejdere skal udfylde, hvis de ønsker et nyt produkt udviklet. Templaten skal beskrive hvad formålet er med produktet, hvornår det skal leveres og hvilke data der skal

Figur 4: Viborg F.F Dasboard (PowerPoint udleveret af Viborg F.F)

anvendes. Derudover er der ikke nogen formelt definerede processer for hvordan de i afdelingerne skal arbejde med data og kommunikere det. Samtidig har man på nuværende tidspunkt kun meget få automatiske kontroller på plads, og kvalitetssikring sker primært gennem manuelt arbejde udført af de få medarbejdere, der arbejder med data til daglig. Alligevel er der sket forbedringer i forhold til tidligere, "… hvor tidligere så var det en manuel proces, der godt kunne tage en halv dag, fordi så skulle man åbne alle filerne. måske printe dem ud. Skrive over i i hvad hedder det i Word måske så det over i Excel" (Bilag 4, D.2: Præsentation af Daniel).

I dag fungerer det bedre i samarbejde med Olga, idet der er bedre mulighed for at bede om den data man skal bruge: "der har vi egentlig gjort det sådan at nu kan de bare komme og bede om det, de har lyst til, og så sørger vi for at datainfrastrukturen er, og så er det egentlig bare tæt samarbejde med dem om, hvordan skal det her se ud, sådan så det viser de ting, de skal bruge." (Bilag 4, D.2: Præsentation af Daniel)

Fremover vil Viborg FF, skulle have fokus på risikostyring, ved at automatisere flere af processerne, proces kontrollere tilgangen i hver afdeling

- Automatisere processer der er manuelle
- Udarbejdelse af datahåndbog
- Automatiske kvalitets tjeks
- Dokumentation i data-log
- Standardiseret tilgang til databehandling
- Ansvarsfordeling i de enkelte afdeling
- Revurdering ML-model (relevans af model)
- Performance reviews af kamp-, spiller- og tilskuerdata

Selvom målet er fuld automatisering, erkender man, at hastighed nogle gange kræver manuelt arbejde, da Daniel f.eks. nævner: "men hvad nu hvis vi introducerer bare et manuelt step, sådan så I lidt hurtigere kan komme i gang øh med at få de ting, I godt kunne tænke jer at vide" (Bilag 4, D.2: Præsentation af Daniel)

(Data Governance Institute u.å.)

A.5 Accountabilities:

Viborg FF efterspørger netop mindre personafhængighed og mere systematisk ansvarsfordeling, hvilket stemmer overens med Data Governance principper om at implementere governance-roller. De er meget bevidste om risikoen ved at være for personafhængig, hvor viden flytter med medarbejderen: "altså du kan have medarbejdere, der sidder ude i… som sidder mega mega dygtige og skifter arbejdsplads, og de tager bare deres computer og flytter med, og så er der bare intet tilbage for virksomheden" (Bilag 4, D.2: Præsentation af Daniel). Målet er derfor klart og tydeligt at: "Hvis jeg ikke er her i morgen, så skal vi kunne arbejde videre" (Bilag 4, D.2: Præsentation af Daniel). Et fremtidigt struktureret opsyn, inklusive KPI'er for datakvalitet, vil styrke klubben overordnet set. På nuværende tidspunkt er der ikke fordelt data governance roller i Viborg FF. Daniel og Olga påtager sig rollerne, men der er ikke klare retningslinjer for, hvem der har hvilket ansvar. Hvilket betyder, at den enkelte afdeling ved ikke, hvem de skal henvende sig til i de enkelte situationer.

- **Data stakeholders:** Afdelingerne (Administration, Sporten, Fremmøde/Billetsalg)
- **Data Stewards:** Daniel, Olga, Anja, Josephine, Praktikanter
- **Custodians:** Olga, Eventii, NFC, Second Spectrum samt andre dataleverandører

Big G inkluderer opgaver og beslutninger så som:

- Oprette datawarehouse
- Data ejerskab strategi
- Data på organisationsniveauet
- Sikring af data (GDPR)
- Rollefordeling
- Daniel fastlægger datastandarder for hele klubben

Little G inkluderer opgaver i det daglige så som:

- Dokumentation af datakilder
- Udarbejdelse og rensning af data
- Procesbeskrivelser
- Standardiserede måder at indtaste data på
- Olga laver validering af data efter kampe
- Palle indtaster kampdagsdata

(Data Governance Institute u.å.)

A.6 Decision Rights

For at kunne analysere Viborg FF's beslutningsprocesser i forhold til deres data, er der lavet et RACI-matrix til at vise hvilke roller, der har forskellige ansvar og deres roller i beslutningerne der bliver taget i organisationen.

RACI-matrix	Direktion	Daniel	Olga	Afdelinger	Eksterne
Datastandarder	I	A/R	C	I	-
Datainfrastruktur	I	A/R	C	I	C
Datakvalitet	I	A	R	C/R	-
Forretningsregler	R	I	C	A/R	-
Analyser & rapportering	I	A	R	C	-
ML-modeller	I	A	R	C	-
Nye dataprojekter	A	R	C	C	I

Figur 5: Raci-Matrix (Egen tilvirkning)

Hvem er del af beslutningsprocessen i Viborg FF, der er listet en række af relevante afdelinger og personer nedenstående som RACI-matrix er udviklet på baggrund af.

Head of Data (Daniel)

- Godkendelse af datastandarder
- Beslutninger om datainfrastruktur
- Dataprojekter

Data Analyst (Olga)

- Tjek datakvalitet
- Analyser + forberedelse af data til ML-modellen/modeller
- Kontaktpunkt for andre afdelinger, fx Palle

Afdelinger (Sport, Administration, Billetter/Fremmøde)

- Forretningsregler / guidelines
- Beslutter hvilke data som er relevant for driften af virksomheden

Eksterne (Eventii, NFC, Second Spectrum)

- Billet data
- Spiller og performance data

Direktion

- Bestyrelse
- CEO/CFO

(Data Governance Institute u.å.) (Figur 8: Organisations Diagram)

A.7 Policies & Rules:

Det er vigtigt for Viborg FF at opsætte et sæt af regler for hvordan dataen skal arbejdes med, det kan hjælpe medarbejderne med at opnå et højere kvalitetsniveau. Det er især vigtigt at undgå de problemer, der opstår, når datastrukturen ændres uden klare retningslinjer, hvilket kan gøre tidligere automatiseringer svære at bruge: "Så alt det man har brugt tid på at automatisere, det er faktisk, jeg vil ikke sige ikke brugbart, men det var i hvert fald svært at sætte i brug [efter datastrukturen blev ændret]" (Bilag 4, D.2: Præsentation af Daniel). Vigtigheden af dette understreges også af nødvendigheden af at kunne overtage hinandens arbejde ved fravær: "det skal være sådan at hvis Olga tager på ferie i 3 uger eller får tilbudt et jo i novo nordisk, at jeg [Daniel] så vil kunne overtage og forstå hendes kode" (Bilag 4, D.3: Præsentation af Olga & Praktikanterne). Fremadrettet kan de undgå problemer og øge deres datamodenhed ved at opsætte et regelsæt.

Regler:

- Kvalitets tjek – Data skal være rene, komplette og opdaterede
- Alle medarbejdere skal kende datakilder og processer
- Sikkerhed & GDPR – Persondata som sæsonkort, abonnement og spillere skal beskyttes
- Tilgængelighed – Data bruges aktivt i beslutninger, fx ved Palle og indkøb af varer
- Ansvarlighed – Sørge for altid at tage Ansvar for opgaverne
- At skabe værdi – Data skal understøtte sportslig og kommerciel performance
- Standard for datakvalitet – Faste valideringsregler (dataformat, manglende værdier, snake_case i kode)
- Retningslinjer for indsamling – Alt data skal tjekkes for GDPR
- Dataopbevaring – Definerede regler for fans, spillere og sponsor.
- Dokumentation – Alle datakilder dokumenteres i en data-log styret af Daniel (Data Governance Institute u.å.)

A.8 Data Governance Process, Tools and communications

Daniel vil bære ansvaret som Data Governance Officer(DGO), hvilket gør ham ansvarlig for at kommunikere nye tiltag inden for datahåndtering ud til de andre afdelinger i organisationen. Det er vigtigt, at han er konkret og onboarder medarbejderne gennem en visuel gennemgang af de nye data regler, da de her krav til data efterspørgsler allerede eksisterer: "Det kræver, vi gør lige nøjagtigt de her ting her. Så det har vi valgt at implementere, det kommer til at påvirke jeres arbejde lige nøjagtigt sådan her. Så I skal gøre sådan her, sådan her og sådan. Og så laver vi en guide til dem. Og så det er sådan her det skal gøres frem" (Bilag 4, D.6:

Interview med Daniel & Praktikanterne), men der mangler konkrete og faste regler, som f.eks. kunne fastgøres i en datahåndbog. Daniel og Olga vil løbende tage målinger af, hvor effektive indgrebene er i realiteten, da den hurtige spredning af data efterspørgsel giver travlhed: "nu er det meget mere. Vi skal faktisk prøve at sørge for, at vi bare kan nå at holde fast i eller Nå de projekter vi har, fordi vi har rigelige opgaver, fordi det spreder sig, fordi man oplever, okay, det jeg har brugt fire timer før hver eneste dag på, det er for ham her løst på to minutter" (Bilag 4, D.6: Interview med Daniel & Praktikanterne). Disse målinger vil blive præsenteret løbende for afdelingerne som en testimoni for fremgang.

For at få en effektiv kommunikation, skal Daniel kunne:

- Formidle programmets værdi og mål
- Skabe balance og forståelse mellem forretning, compliance og teknik
- Beskrive processer og resultater
- Dokumentere beslutninger, ansvar og metrics
- Formulere politikker og gældende regler
- Kommunikere tekniske emner til ikke-tekniske kolleger
- Indsamle succeshistorier og behov fra interessenter

VFF arbejder i dag med en række systemer, som alle generer nye data dagligt. Den sportslige afdeling bruger PowerApps til at registrere træningsdata, som efterfølgende analyseres og visualiseres i PowerBI. Afdelingen indsamler også spillernes præstationsdata under træning og kampe ved at anvende GPS-veste, som giver indsigt i spillernes belastning og løbemønstre.

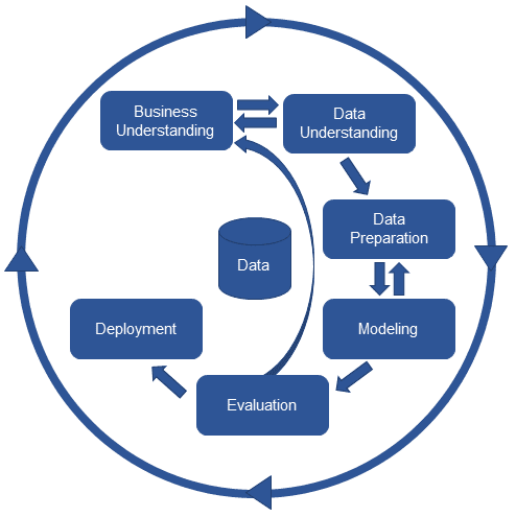
På den kommercielle side anvender salgsafdelingen et CRM-system til at håndtere sponsorater, aftaler, indtjening og kundekontakt. Marketingafdelingen arbejder med data fra sociale medier, salg af merchandise og VFF app'en, hvor man kan følge fansenes adfærd. På nuværende tidspunkt lagres data primært i Azure og Sharepoint, men i fremtiden ønsker man at udvikle sit eget datawarehouse, som kan samle data fra mange forskellige systemer og lagre det.

Dataafdelingens opgave er at fungere som bindeled mellem forretning og data. Afdelingen beskriver og standardisere processer, udarbejder politikker og fastsætter regler, der sikre en ensartet håndtering af data på tværs af organisationen. Kommunikation af beslutninger vedrørende data governance formidles primært gennem præsentationer på møder eller mail korrespondance via outlook.

- Indsamling af data – Palle indtaster kampdagsdata og Olga validerer
- Olga renser og strukturer data i standardformat
- Lagring af data - Alt data gemmes i fælles database/data-log
- Analyser – Data bruges til ML-modellen, kommercielle analyser og sportslig performance
- Rapportering – Resultater leveres til ledelse, marketing og sportsafdeling
- Feedback – Afdelinger giver feedback til Daniel → processen opdateres efter behov

(Data Governance Institute u.å.)

A.9 Data Governance Work Program



Figur 6: CRISP-Model. (Udleveret af Bjarne)

Ved implementering af en datadreven arbejdskultur, hvor de enkelte afdelinger følger datahåndbogen og Daniel indtræder som Data Governance Officer, vil næste skridt være oprettelse af workstreams. I en virksomhed der arbejder med data på en projektbaseret basis handler workstreams om at hvert projekt har en livscyklus. Daniels indsats som Data Governance Officer skal være med til at flytte organisationen fra differentiationsperspektiv, til integrationsperspektiv og hermed nedbryde siloer og forskellige grundlæggende antagelser af hvad data er "vi skal blive lidt skarpere på, at vi snakker om det samme, når vi snakker om data i en, hvad kan man sige, i en i en virksomhed" (Bilag 4, D.2: Præsentation af Daniel). Som Data Governance Officer skal Daniel skelne mellem om projektet har "bløde" resultater, mens andre skal måles med SMART-tankegangen.

- Specific (konkrete)
- Measurable (målbare)
- Actionable (handlingsorienterede)
- Relevant (relevante)
- Timely (rettidige)

Data Governance-programmerne skal være realistiske i deres scope og ressourcefordeling. De skal kunne levere værdi der hvor det er lovet, og prioritere de rigtige opgaver i de rigtige workstreams. Implementering af data governance work program for Viborg FF vil medføre værdien af den tilgængelige data, medvirker til at forbedre ressourceplanlægning og hermed mindske madspild. Viborg FF får en øget fitness-for-use for deres data ved at revurdere deres modeller og dets outputs.

(Data Governance Institute u.å.)

A.10 Participants

Participants er det som Viborg FF i fremtiden burde arbejde imod at dele rollerne ud i organisationen for at skabe et bedre arbejdsmiljø og strømline deres processer gennem standardformater og skarpe roller, hvor at ledere og medarbejdere nemt kan effektivisere den daglige drift for at opnå bedre resultater og opnå mere.

- **Data Governance Officer(DGO):** Daniel
- **Data stakeholders:** Afdelingerne (Administration, Sporten, Fremmøde/Billetsalg)
- **Data Stewards:** Olga, Anja, Josephine, Praktikanter
- **Custodians:** Olga, Eventii, NFC, Second Spectrum samt andre dataleverandører

Big G inkluderer opgaver og beslutninger så som:

- Oprette datawarehouse
- Data ejerskab strategi
- Data på organisationsniveauet
- Sikring af data (GDPR)
- Rollefordeling
- Daniel fastlægger datastandarder for hele klubben

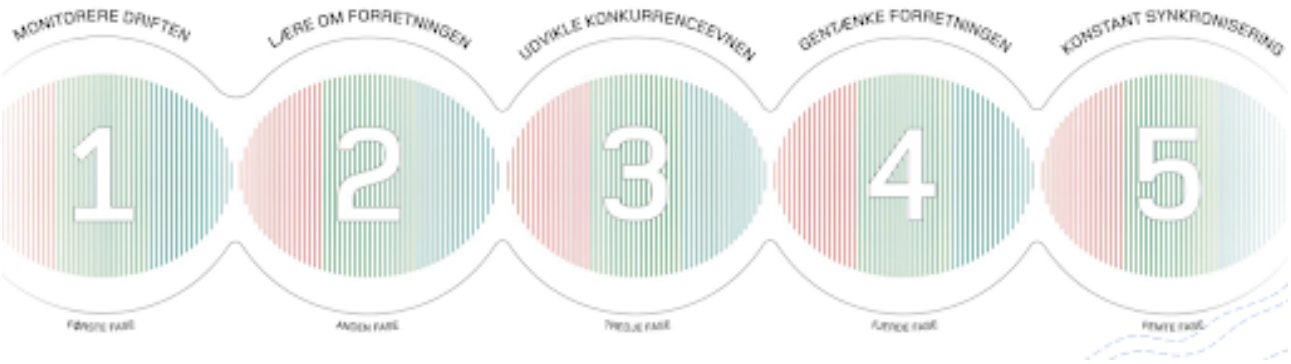
Little G inkluderer opgaver i det daglige så som:

- Dokumentation af datakilder
- Udarbejdelse og rensning af data
- Procesbeskrivelser
- Standardiserede måder at indtaste data på
- Olga laver validering af data efter kampe
- Palle indtaster kampdagsdata

Ved at give klare opgaver til Data Stewards/Custodians som kan arbejde med Little G (de daglige opgaver), vil det gøre arbejdsdagene nemmere for medarbejdere, men også skabe klare resultater og løse nye problemer hurtigere. Yderligere kan Big G (strategiske beslutninger) være med til at give VFFs ledelse bedre mulighed for at tage beslutninger om hvordan nye arbejdsmetoder kan effektiviseres, samt skabe klare retningslinjer for medarbejderne i Little G. Et eksempel kunne være en ændring i lovgivningen i forhold til GDPR, som Daniel samt direktionen tager en beslutning om at ændre arbejdsprocesserne blandt andet ved enten spillerinformation, sæsonkort eller lignende. Det kunne også inkludere ændring af datalagring. (Data Governance Institute u.å.)

B Bilag 2: Datamodenhed

På baggrund af den udarbejdede Data Governance analyse, samt de interviews der blev foretaget under besøget ved Viborg FF, vurderes deres nuværende data modenhed til at ligge i det accelererende trin i fase 2. VFF viser tydelige tegn på at være på det accelererede trin, da data langsomt er blevet fastslået i virksomhedens kultur. Der arbejdes med at anvende data til at analysere og forbedre sportslige og kommercielle resultater, som kan medvirke til at reducere madspild, monitorere spillernes belastning og forbedre præstationer under kampene. VFF viser dog også tegn på at være på vej mod fase 3, da data ses som en konkurrenceparameter, hvor man ser sig selv som en af landets førende klubber inden for datamodenhed. Der er oprettet en dataafdeling, som består af data stewards og custodians, hvor Olga som data steward, allerede har klare opgaver om at rense, klargøre og vedligeholde data. Placeringen af fase 2 er baseret på en vurdering af Viborg FF's datastruktur, roller og processer der bruges dagligt. (Kølsen, Nielsen, og Bækby 2017)



Figur 7: Datamodenheds faser (Kølsen, Nielsen, og Bækby 2017)

VFF viser flere tegn på at være forbi fase 1 (Monitorere Driften). Data er ikke længere spredt uden styring, der findes faste systemer til både den sportslige og kommercielle data, blandt andet gennem CRM, Eventii og BI-dashboards (Kølsen, Nielsen, og Bækby 2017) . I interviewet nævner Palle at han har efterspurgt data på at kunne forbedre indkøbet af mad og drikke til hjemmekampe, samt resourceplanlægge bedre. Det viser at andre afdelinger i virksomheden begynder at efterspørge bedre datagrundlag og ikke befinder sig i den indledende fase, hvor at dataarbejde er tilfældigt og mangler retning. Daniel udtrykker også, at næste skridt for Viborg FF's dataudvikling skal være oprettelse af et data warehouse.

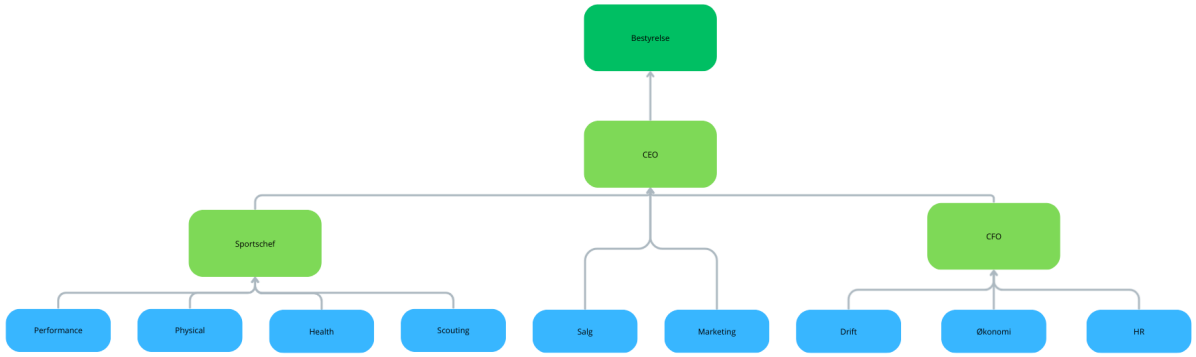
Yderligere er nogle grundlæggende kendetegn for at organisationen er i fase 2 blandt andet at datahåndteringen stadig er præget af mange manuelle arbejdsgange, uformelle roller og mangel på faste processer. (Kølsen, Nielsen, og Bækby 2017) Datakvaliteten kontrolleres af enkelte personer som Daniel og Olga, samt praktikanter. Daniel nævner at "jeg oplever generelt en rigtig stor nysgerrighed på, hvad vi kan få ud af vores data" (Bilag 4, D.2: Præsentation af Daniel), hvilket viser en øget interesse fra andre afdelinger og ledelsen.

Viborg FF har vist en stærk bevægelse fra at overvåge driften (Fase 1) til at bruge data til tværgående læring (Fase 2) med en klar ambition om at udvikle konkurrenceevnen (Fase 3). For at kunne bevæge sig over i fase 3, skal Viborg FF implementere data governance (Bilag 1: Data Governance). Implementering af data governance

vil medføre nedbrydning af siloer internt i klubben, på tværs af afdelingerne. Dette vil medføre en fælles forståelse for data, it og forretningsudvikling, samt bedre tværfaglig samarbejde på tværs af klubben. Viborg FF viser indikationer på at de er på vej imod fase 3, Daniel nævner "Det har jo ingen betydning altså for hvor høje spillere er til noget som helst i vores forretning […]. Det vi skal være gode til det er, når vi nu har den data, så skal vi sørge for, at vi ikke bruger alt for lang tid på at undersøge alle mulige ting. Vi skal bruge data, der kan træffe beslutninger" (Bilag 4, D.2: Præsentation af Daniel). Derved viser Daniel et stærkt syn på, hvilken retning Viborg FF arbejder imod, samt hvilket formål de har med at bruge dataen.

Vi vurderer ud fra Data Governance- og datamodenheds analysen at den bedste fremgangsmåde fremadrettet for Viborg FF er at udvide dataafdelingen med en til to medarbejdere til at hjælpe med at automatisere og udvikle processer som standardisere deres arbejde, Daniel udtaler "hvad er det for nogle udfordringer vi egentlig har i virksomheden […]. Vi har de her processer. Det tager faktisk fire timer for en medarbejder at gøre hver dag. Vi har tre af de her medarbejdere. Det er 12 timer om dagen, som bliver brugt på noget, som er fuldstændig manuelt. Det skal fungere automatisk." (Bilag 4, D.6: Interview med Daniel & Praktikanterne). Ud fra hans kommentarer, vurderes det at en mangel på arbejdskraft spiller en stor rolle for hvad de kan nå i afdelingen i det daglige. (Kølsen, Nielsen, og Bækby 2017)

C Bilag 3: Organisations diagram



Figur 8: Organisations diagram (Viborg F.F. 2019, Egen tilvirkning)

D Bilag 4: Tematiserede interviews

D.1 Temaer

- Datakultur
- Dataarbejde
- Data governance og struktur
- Kapacitetsudnyttelse
- Billetaktivering og fremmøde
- Økonomi
- Systemlandskab
- Placeholder

D.2 Præsentation af Daniel

Data governance og struktur

"I dag så har jeg [Daniel] en anden titel Head of Data and Technology,"

"jeg [Daniel] varetager rigtig rigtig mange forskellige opgaver, men hovedopgave det er hvordan vi indsamler opbevarer og bruger data på tværs"

"administrationen, det er alle dem, der var tager opgaver, som er udenfor fodboldbanen."

"Den anden halvdel af virksomheden tager sig kerneopgaven, som jeg plejer at sige. Det er det, der har noget at gøre med fodboldspillet."

"opdeler det sådan i fire afdelinger. Vi har en driftsafdeling. Det er en tre fire mennesker som sørger for at der er mad i hovedet. Der bliver solgt billetter. Der er nok personale på Alle de ting der skal til for at facilitere en fodboldkamp."

"Salg. De står for aktiveringssponsorater. Det vil sige de taler med virksomheder i forhold til hvordan kan vi finde et eller andet partnerskab, hvor vi kan tilbyde noget. Ofte er det, vi kan tilbyde en unik platform i I forhold til at vi er den mest eksponerede virksomhed i kommunen"

"Så har vi siden af en afdeling, som hedder marketing og kommunikation"

"Der er tre fuldtidssælgere og en salgchef. I marketing og kommunikation der sidder to content createre."

"Så er der en marketingsansvarlig. Så er der en ansvarlig for vores LinkedIn"

"Økonomi er en fuldtidsmedarbejder, som hedder Pia, hun var dernede. Og så vores CFO, vores økonomidirektør"

"Men det det [Palle] handler om, det er driften af boder og billetter salg"

"Den første afdeling jeg vil komme ind på det er den vi kalder performance. Og i mange fodboldklubber så er performance noget med fysisk data, men den kalder vi physical hos os."

Billetaktivering og fremmøde

"Når I kigger op på stadion, det her det er et billede af stadion, og I kan se, der har solgt nogle sponsorater til nogle virksomheder. Jamen, så Så skal vi jo have nogle sælgere ud og aktivere de her virksomheder, som ligesom kan sørge for, at der bliver lavet nogle partnerskaber."

"vi har nogle sælgere, der tager ud til virksomheder og sørger for at vi er stærkt repræsenteret af det, hvad kan man sige, virksomhedsnetværk der er i Viborg."

Kapacitetsudnyttelse

"Så når man hænger et skilt op på stadion, så bliver det set af rigtig rigtig mange mennesker. Så vi har en platform vi kan sælge på."

"det vi bedst muligt kan tilbyde, det er en platform, hvor man kan blive set og hørt, og man kan skabe forbindelser på tværs af forskellige virksomheder."

"vi laver forskellige netværker, som man, hvis man er en tømmervirksomhed, så kan man møde en mailervirksomhed, som kan møde en bankvirksomhed, så vi kan sætte folk sammen."

"vi lige præcis skal målrette vores kommunikation hen imod de enkelte fans. På sigt så er planen sådan set at vi godt kunne tænke os at blive endnu skarpere på det"

"så når vi taler med sponsorer at vi kan være meget tydelige omkring okay du er en bankvirksomhed. Vi har lige nøjagtig de her fans og som i kan kommunikere til igennem os. Det er et mål vi har"

"Så er der data på merchandise. Altså hvad er det for noget vi sælger? Hvornår er det vi sælger det? Hvad er det vi skal være opmærksomme på og så videre?"

Dataarbejde

"På datasiden, så har vi et CM-system øh hvor vi registrerer aftaler, holder styr på produkter og indtjening og så videre"

"det er her data tit bliver født hos os"

"når der er en ny partner, som starter hos os, så bliver de lagt ind i systemet"

"Igennem den [appen] så får vi også indsamlet en masse kan man sige stamdata på øh fans"

"I dag der i samarbejde med Olga, der har vi egentlig gjort det sådan at nu kan de bare komme og bede om det, de har lyst til, og så sørger vi for at datainfrastrukturen er, og så er det egentlig bare tæt samarbejde med dem om, hvordan skal det her se ud, sådan så det viser de ting, de skal bruge. Så nu kan de bare skrive til os, jeg skal bruge øh vores KPI'er på de sidste 10 kampe, så kan vi give det til dem, hvor tidligere så var det en manuel proces, der godt kunne tage en halv dag, fordi så skulle man åbne alle filerne. måske printe dem ud. Skrive over i i hvad hedder det i Word måske så det over i Excel få koderne til at virke i Excel"

"vores opgave er lidt det samme. Indsamle opgaver og præsentere sådan så at Dem der skal bruge det"

"Dem der skal bruge det, hvis de fysisk trænere i stedet for de skal sidde i Excel og downloade og smide over og alle mulige forskellige ting, så går de bare ind og kigger, og så får de egentlig bare data præsenteret sådan så at deres opgave bliver at kigge på det færdige produkt og ikke lave det, fordi det er ikke deres kompetencer."

"men der kan jo være enormt meget viden, som bliver lavet i de enkelte afdelinger, som kan bruges på tværs. Et eksempel på det er, at vi i samarbejde med den fysiske sektor havde rigtig store udfordringer med at hjælpe dem, fordi at de mente ikke at de kunne kalde spillerne, hvad de har lyst til. Så de kaldte spillerne deres nummer og så deres efternavn. Så når spillerne de skiftede numre, det gjorde de nogle gange, så var det forskellige spillere i deres datasæt, og så kunne de slet ikke finde ud af, hvad det egentlig var. Så det vi gjorde, det var, vi lavede unikke ID'er til hver eneste spiller på tværs af hele Viborg FF."

"Men jeg plejer at sammenligne det med når jeg skal rengøre min lejlighed så hvis jeg skal gøre den sådan 90% rent så tager det måske en time en halv time at gøre den ring hvis jeg skal gøre den fuldstændig ring. Altså der må ikke være noget som helst i lejligheden. Øh jeg har et fem måneder gammel barn så det er faktisk fuldstændig umuligt. Men lad os bare sige det to en dag. Øhm. Jamen så så tager det proportionelt sindssyg lang tid at lave det sidste. Og det siger vi også med fuldautomatiske løsninger. Hvis vi skal lave en fuldautomatisk løsning, så vil I bare opleve, at det kommer til at tage sindssygt lang tid, for I har det øh potentielt, hvis dataen ikke er der. Øhm og det er bare vigtigt at få sagt, fordi når de så siger, jamen det skal virke fuld automatisk, så er det lidt nemmere for os at sige, jamen så kommer det bare til at tage længere tid på det"

"men hvad nu hvis vi introducerer bare et manuelt step, sådan så I lidt hurtigere kan komme i gang øh med at få de ting, I godt kunne tænke jer at vide. Det kunne faktisk eksempel være at vi merchandingsafdelingen har et system der tillader CSV eksport jamen så kan vi downloade det lægge det i den rigtige mappe og så opdatere PowerBI så har vi faktisk det vi skal bruge."

"for at vise dem at der er ligesom tre ting, ikke? Altså jeg har datakilden, det kunne være en ekstern leverandør. Vi har et eller andet der skal trække data fra datakilden over til det er som regel das gerne væk. Okay. Hvad er det, der skal trække det her data? Er det mig der skal gå ind og downloade det og lægge det i den rigtige mappe og så trykke opdatere det ligesom det der skal ske? Skal vi have sat noget kode op der trækker det automatisk? Hvad er det, der skal på en eller anden måde få data fra den her datakilde flyttet over til den præsentation, de gerne vil se? Og det er mere for sådan at sætte hele tiden billede på, at der er faktisk forskellige dele af det arbejde, vi laver"

"jamen fra at Claus i merchandise siger, jeg vil gerne have et dashboard, det skal se sådan her ud til at vi ender med at have det dashboard. Der kan godt gå rigtig lang tid, og der er rigtig lang tid, hvor han oplever at jamen der sker jo ikke noget. De sidder jo bare der med deres computer og sådan noget. Siger jamen det er helt almindeligt at fra når vi kommer til dig og siger, er det sådan her det skal se ud til vi tester nogle forskellige ting, der er vi stort set færdige. Men Men indtil vi er der, der kan det godt gå rigtig lang tid, før du oplever der rent faktisk kommer nu. Fordi hvis jeg kigger på datastrukturen, der er fejl i datastrukturen. Hvorfor ser det sådan her ud? Vi skal måske have fat i leverandøren. Vi skal have fat i Claus igen. Alle de her ting her. Okay. Og lad os så sige, at ø man ansatte en praktikant, som hed Olga. Hun gik i gang, og hun kæmpede med det der

billetdato. Hun fik det sku til at virke, og så ændrede man lige datastrukturen. Så tog man fat i leverandøren og sagde, det ville ville ikke være smartere, hvis man gjorde det her. Og så kommer Olga tilbage, og så lige pludselig så er datastrukturen fuldstændig ændret. Så alt det man har brugt tid på at automatisere, det er faktisk, jeg vil ikke sige ikke brugbart, men det var i hvert fald svært at sætte i brug" \ "vi skal kigge lidt mere på det her data warehouse, fordi vi kan godt over tid øh få tingene til til at fungere med den størrelse vi har nu med at gøre ting i Sharepoint, gøre ting med R, gøre ting med PowerBI, det kan vi godt få til at fungere. Men der kommer et tidspunkt om ikke sådan alt for længe, hvor at vi har så mange, hvad kan man sige, projekter i gang, at det er svært at gøre uden at få det næste lag på. Så skal vi have en større afdeling"

Datakultur

"jeg oplever generelt en rigtig stor nysgerrighed på, hvad vi kan få ud af vores data. Jeg synes, der hvor vi oplever øh udfordringer, det er forstå, jeg kommer ind på det lige om lidt også, forståelsen af hvor lang tid forskellige dele arbejdet med data kan tage, hvor at hvis du ikke har en datamæssig baggrund, kan du godt tro, at det tager jo ikke ret lang tid"

"at der har været rigtig meget, hvad kan man sige, uddannelse, afdelinger, uddannelse af medarbejdere i forhold til, hvad er det, der tager tid, når vi når vi går i gang? Hvad er det så, I kommer til at opleve, hvornår er det, det ikke virker og sådan nogle ting der, det har der været en del af, fordi vi er en virksomhed som er meget sådan autodidakt oplært."

"vi kigger meget på, er du medarbejdereafhængig, eller er du leverandør afhængig? Og vi var lidt begge dele."

"altså du kan have medarbejdere, der sidder ude i, nu kommer jeg til Silo også lidt senere, som sidder mega mega dygtige og skifter arbejdsplads, og de tager bare deres computer og flytter med, og så er der bare intet tilbage for virksomheden. Altså sådan i der er ikke noget data, der er ikke nogen, der ved, hvad der er blevet lavet, så kommer der bare en ny en med en PC og begynder at arbejde. Og du har ingen anelse om, om det er det samme, om det er forskelligt, du har ingen chance. Og det vi så gjorde det sige, vi skal ikke være medarbejderafhængige. Hvis jeg ikke er her i morgen, så skal vi kunne arbejde videre. Hvis vi har en leverandør, der skruer prisen på højt op, så skal vi kunne arbejde videre"

"Men det der godt kan ske det er at afdelingerne nogle gange godt kan have lidt svært ved at have tid til at snakke sammen med hinanden. Og derfor så er en del af vores opgave også at sørge for de ting som vi får implementeret som kan hjælpe på tværs det også bliver kommunikeret på tværs"

"vi skal blive lidt skarpere på, at vi snakker om det samme, når vi snakker om data i en, hvad kan man sige, i en i en virksomhed Vi gør det sådan, at vi siger de første fire spørgsmål Det er noget de skal besvare. Det sidste det skal vi besvare. Det vil sige hvor hvor skal det fremvises? Skal det laves i en PDF? Skal de have det på mail? Skal der laves en platform? Skal der laves PowerB? Hvor er det skal fremvises til dem?"

"Her har vi tit udfordring med at folk ikke rigtig er klar over hvad det egentlig er de gerne vil have vist. De vil bare gerne have vist data."

“Nu er det Claus er ansvarlig for merchandise, men han har også en medarbejder Simon. Jamen skal de to se det samme eller skal de se to forskellige ting? Det kan jo godt være han som leder har behov for at se nogle ting, som Simon ikke skal se det er vi simpelthen gjort for at være sådan meget tydelige på, hvad er det vi egentlig har brug for, og hvornår er det, at vi ikke længere har brug for feedback til, hvor i Excel det skal ligge henne.”

“Så kigger vi på, jamen det der data i troede, der var der på spiller, tror jeg, men det det er der jo slet ikke. Altså vi har slet ikke noget data på det. Jamen, okay, så er det retur, og så begynder vi at stille krav til, I vil gerne have det her, men vi har jo ikke data på det. Vi tror på, vi kan gøre det på den her måde. Er det noget, vi skal gå videre med, eller? Ja, fedt. Vi kan godt gå videre. Vi har faktisk lige præcis data vi skal bruge”

“alle vil gerne have fuld automatiske løsninger.”

“det der med at være en autodidakt virksomhed, datastruktur kan jo være mange dyr i skoven. Og der prøver vi bare at være sådan lidt skrappe på, hvad er det egentlig”

“For at få en dialog omkring, hvorfor er det, det er så vigtigt at have det her ark her? Fordi det her ark indeholder alt den information, vi skal bruge, hvor at hernede der er ID lige pludselig flyttet op som en kolonne. Jamen, hvorfor var det et problem? Det er da ikke noget problem. Jeg er der bare lige rykker rundt på det, og så lige pludselig oplever de, at de ikke får deres ting” \ “Hvad for nogle filformater, der plejer jeg bare at sige, de er faktisk ligeglade. Bare vi sørger for at holde sådan rimelig ens, sådan så vi ikke nogen uger har øh billeder, og nogle uger har vi CSV-filer”

“så har de lige tænkt, jeg retter lige på det, og så smider de det op i den mappe i stedet for de bare tog rådataen og lagde op” \ “Fordi hvis jeg nu hver eneste dag spillerne kom ind, målte deres højde, så havde jeg jo det bedste data i verden på, hvor høje var spillere. Er det så har været gode til at arbejde med data? Det har jo ingen betydning altså for hvor høje spillere er til noget som helst i vores forretning. Så derfor så er man bare nødt til at være meget klar på, hvad er det vi skal være gode til? Og det vi skal være gode til Det er, når vi nu har den data, så skal vi sørge for, at vi ikke øh bruger alt for lang tid på at undersøge alle mulige ting. Vi skal bruge data, der kan træffe beslutninger”

D.3 Præsentation af Olga og praktikanterne

Datakultur

” Det her er så et eksempel på en testen hvor jeg. Imens jeg lige har været her, er der ikke blevet lavet den her test. Da jeg ikke har været med ude og observere, hvilken viste sig at være ret vigtigt. Til gengælde har jeg fået en god beskrivelse. Men det er heller ikke dem der skal tilpasse sig til min løsning, det er mig der skal løse det.”

”Under praktikken har det været meget projektbaseret. Jeg har haft et bruttodataprojekt, som jeg nævnte sidst, vi var ude ved jer og præsentere. Jeg har fået udleveret en masse data fra stadion – koderne på stadion – fra deres poster, i alle mulige formater: PDF, Excel-filer og alt muligt.”

Dataarbejde

“Så har jeg haft et andet projekt, som jeg bare har kaldt spillerstatistik. Det handler om benchmarking på køb, salg og leje af spillere i Superligaen generelt fra 2021 og frem. Det er stadig et igangværende projekt. Det har primært involveret webscraping fra Wikipedia og fra Superligahold, der har spillet forskellige år. Og så har vi haft Transfermarkt, hvor man kan se, hvilke spillere der har været på hvilke hold.”

“Helt i starten blev vi også præsenteret for Power BI og Power Apps. Jeg har mest brugt Power BI til at lave nogle visualiseringer af den strukturerede data. Anja har brugt Power Apps rigtig meget.”

“Jeg er også begyndt på en dataanalyse i R. I har hørt Daniel nævne, at nogle sidder i den sportslige afdeling og laver KPI’er for, hvad de skal kigge på. Jeg har fået GPS-data og andre data og selv fundet nogle, og forsøgt at gå til det med mine nysgerrige øjne udefra – jeg går ikke vildt meget op i fodbold, men jeg ved mere nu.”

Data governance og struktur

” Der også nogle spørgsmål som jeg er kommet med, for jeg synes også at det kunne være interessant. Men jeg vil også sige at jeg er blevet meget positiv overrasket over den sportslige afdeling, som er dem jeg har arbejdet mest med. De ved rigtig mange ting og vil også gerne have mange ting, men det skal også bare være muligt. Vi skl nå dertil før vi kan lave det”

” Vi er stadig der hvor vi kigger på hvad PowerBi kan og der har de to her (det to praktikanter) været til meget hjælp”

“Jeg skulle finde ud af, hvordan vi kunne få struktureret den data, så vi kunne visualisere den.”

“Vi har også arbejdet med mappestruktur, organisering og dokumentation af koden. Hvordan man bedst muligt gør det, så man kan finde tilbage til, hvad man har kodet og hvorfor. Og hvis jeg skal overdrage mine opgaver, når praktikken slutter, kan andre følge med i projektet.”

“Den starter lidt ’der’. Vi har kigget på noget kritisk tænkning. Det er ikke fordi, vi skal sidde og forklare, hvad alt det her er – for det kan Power BI normalt gøre ved at trykke på knappen. Men vi har lavet noget tekst-fortolkning af, hvad vi mener, det er.”

“Jeg lavede en PowerApp, men fandt ud af, at formlen i Excel ikke virkede – så jeg har lavet formlen direkte herinde. Det betyder, at hvis man ændrer noget i Excel-arket, så opdateres det. Jeg har lavet nogle tests, og der er et par småting at gå igennem, men ellers virker det fint. Det her er jo i bund og grund noget, vi skal have 2–3 nye egentlige medarbejdere til at sidde med, og det er også en måde at se, hvor vi skal sætte ind i fremtiden.”

Billetaktivering og fremmøde

“Det, Daniel bad mig om, var at lave en app til salg. Når vi har kamp, så kan de se, hvad der bliver solgt hvor, og hvornår der skal sættes i gang med at producere øl eller popcorn. Altså: hvor meget bliver der egentlig solgt?”

Systemlandskab

” Godt spørgsmål, det skal være sådan at hvis Olga tager på ferie i 3 uger eller får tilbudt et jo i novo nordisk, at jeg så vil kunne overtage og forstå hendes kode. Vi koder forskelligt, men i PowerBi er det nemmere at overdrage,

så det er klart fremtiden"

D.4 Præsentation af Palle

Datakultur Dataarbejde

"Hvis vi kigger på tilsku og udviklingen, jamen ø så kan I se her at 19 øh I første division der var vi en i ar et et tilskuertal på 3.000. Så dykker det selvfølgelig lige her lidt med corona. Men ellers så kan I se hvor hvor stejl en kurve det har gået det har gået opad."

"Øhm jamen så konkrete data her Olga, der skal du nok hjælpe mig lidt, men altså vi, hvad skal man sige? Jeg har næsten lavet den samme måde at opgøre tilskuertallet på alle 20 år, er det ikke sådan næsten rigtigt? Men men dybden i dataen er blevet en anden efter at, hvad skal man sige, vores billetsystem er blevet så detaljeret som det er nu. Så hvad har vi fire år eller sådan noget, hvor vi har den her detaljeringsgrad og sådan noget. Ja. Ja. Og det du sagde data var ok, ikke?"

"Øh, en A-kamp, det er altså en A-kamp, det er de fire, det er de fire store. Det er, hvad er det hedder, Brøndby, det er AGF, det er Midtjylland, og det er FC København. Øh, en A-kamp, det er meget kategoriseret som noget, hvor vi ikke hvor vi hvor vi øh hvor vi kun via markedsføring på kampen kan skabe grundlaget for kampen. Hvorimod en C-kamp, det er det er en kamp, hvad skal man sige, hvor vi er opmærksomme på, hvis vi skal nå et tilskuertal, som er vores måltæ, så skal der ske aktivering."

"Øh i søndags der spiller vi i FC i Parken. Der er en af mine kollegaer Alexander, som står for brodersalg på stadion, han er ovre og følge deres, hvad er det det hedder? Ansvarlig for food andage derovre."

"Og og det det er helt helt ned på detaljen. Altså hvad hvad mål skal de have, hvad omsætning skal de have på den kamp, hvad har de i gennemsnitlig salg per tilskuer, hvad er deres prioriteringer inden for båerne og sådan noget. Der er ekstremt stor åbenhed, fordi at jeg tror generelt så kigger klubberne på at jo bedre vi alle sammen bliver, desto større mulighed er der for at løfte grundlaget for tilskuere i ø i Danmark."

"men altså nu skal jeg passe på, hvad det sådan er, men altså noget noget af det, vi oplevet, når vi arbejder med, det er, at på en måde på en ting så har vi rigtig meget data. Men vi mangler nogle gange noget specifikt på en kamp. Altså for eksempel Vi spiller, vi har 16 kampe om året, men vi møder måske kun holdet en gang, så hvis måske to, så hvis nu at vi har spillet fire sæsoner, så kan det rent faktisk være, du kun, selvom det er 4 gange 16, så kan det være du kun har data fra fire gange fire kampe. Og der kan være helt vildt stor forskel på, om det er maj måned, september måned, december måned, om det er klokken to, klokken 4, klokken seks. Så øh, men jeg har store forventninger. rammer."

Kapacitetsudnyttelse

"Og vi har en kapacitet på 10.000."

"Det har vi brugt ekstremt meget tid på i forhold til vores brug af data øhm og udvikle hele elektroniske platform, udvikle, hvad skal man sige, brugen af af de data vi opsamler kører på kampene ø brugsgrader og hvad er det

det hedder omsætningstal i broderne på mange forskellige fonter. Så så der forsøger vi der forsøger vi at tage nogle skridt og har brugt det til at komme videre med os."

"Så det er sådan nogle ting, vi prøver at bruge det på. Øhm, det samme øh med sæsonkort i ø 1819 der var der var sæsonkort, det var den traditionelle gamle ting med nærmest et plastikkort. som man fik udleveret og man købte på et årskort. Øhm, ret hurtigt så begynder der at ske noget med at ø at man får indført abonnementer på samme måde som Netflix og alt muligt andet."

"Så i dag der er vi 4100 abonnenter på på hvad er det det hedder omkring stadion øh og I kan se der i 1819 der var vi 504 sæsonkortlige sæsonkort holdere så en helt eksklusiv udvikling."

"Og så kan man se jamen så tager man fat i det og begynder at kommunikere med det stiller krav om og hvad gør i fordi for eksempel erhvervslubben, de har alle sammen adgang til spisning. Det er sådan ret nederen at regne med der kommer 800 mennesker og købe mad ind til 800 mennesker så kommer der kun 600. Så øh så det er sådan det er sådan ret tydeligt at det har man taget fat i at det det durer ikke det der. Og der er jo en masse gode historier med madspild og alt muligt andet. som man kan putte på sådan en kommunikation."

"Og i dag noget af det, vi forsøger på derov, det er at putte putte menuer med ned oven i oven i hvad det hedder oven i billetprisen. Så det vil sige når du køber et et sæsonkort abonnement jamen så har du betalt for billetten plus du har betalt for stadionplatten. Jamen mit mål for den største altså vi vi har jo vi har nogle vi har nogle KPI mål. Øh vi har selvfølgelig på et tilskuertal. Vi har noget vi skal omsætte billetmæssigt."

"Vi øh vi er i en situation at øh de folk der arbejder hos os de er som regel nødt til til at have to måske tre jobs, fordi vi har 16 events om året."

"Altså der er markant forskel på at skal servicere 8500 eller skal servicere 6500. Så det er sådan det er sådan meget den forberedelsestid der er der er i det at det er den der er vigtig for os."

Billetaktivering og fremmøde

"Så var der det her med med brugsgrader. Øhm. Her der har vi sådan vores forskellige produkter. Vi har vores billetsalg. Det ligger sådan rimelig pænt med 92% benyttelse."

"Øhm, og vi går ind og kigger på prisstrukturen omkring det. Så vi vi går klart ind og siger, okay, Køber i abonnement, så kan I spare nogle penge."

"Så i dag der Min rolle den går 100% på kampeviklingen. Så når vi har kampe inde på stadionne, så er det mit overordnet ansvar, at hele rammen omkring kampen, den fungerer."

"Øh, nu snakker vi så om gennemsnit, der hvad bruger sådan en øh en juridisk tekstur har det har man tal på,hvad de bruger, når de er inde."

Systemlandskab

"Øhm i forhold til i forhold til hele den her udvikling på datasiden i forhold til de elektroniske, hvad er det det hedder, elektroniske muligheder. Jamen så øh var vi det første stadion i Danmark, der gik over og havde

NFC-sanning. Øh så det vil sige alle vores alle vores, hvad er det det hedder kort ø billetter ø har i dag NFC indarbejdet. Og det betyder jo at når man går ind og skal ind på stadion, så kan man bare lægge sin telefon op på på hvad er det det hedder på møllen, og så er der kommunikation med det"

"Den normale scanning med QR-kode og alt sådan noget ikke fungerer, men det er en kæmpe fordel med NFC scanningen, fordi at mange har har, hvad skal man sige, nedtonet med lyset på på sin skærm, eller der er nok også et par jeres skærme, der er smadret og sådan nogle ting, men der der er der bare ikke noget i forhold til NFC- scanningen, og vi kan se det øh ekstremt meget på vores indgangsdata. Øh hvad det hedder, vi har aldrig kø."

Placeholder

"Ja. Hvert år der der bliver der lavet i Superligaen, der blev lavet sådan en Superliga surervey, hvor hvor klubber bliver altså hvor tilskuerne til klubberne bliver spurgt, hvor man kan man skal man sige måle sig op mod de andre klubber. Så benchmark med dem der. Udover det her så har vi også kørt nogle, hvad er det hedder tilskuerundersøgelser. For tror jeg det to halvanden sæson siden eller sådan noget, hvor vi prøvede at arbejde med det. De der tilskuundersøgelse, det er noget vi gerne vil gøre lidt mere i. Og derudover så så har vi også undersøgelse vores ø hvad er det det hedder hovedpersonale og sådan noget der sådan giver feedback tilbage på hvordan de ligger."

D.5 Interview med Palle og Olga

Datakultur

"Så altså det det vi skal det er at at Olga hun får bygget alle de her rapporteringsværktøjer således at at dataen bare bliver puttet i en en silo der og så kører det. Så der er vi i gang med ret store projekter lige på det."

"For eksempel Palle han er utrolig god til at øh hvad hedder for det første sådan samle på data og så strukturere så alle de der data som ikke ser men de ser ud så kan man tage det og så begynder at arbejde med det."

"Jeg har også data fra evente som er struktureret så det vil sige som jeg sagde at jeg laver nogle transformationer i R hvor jeg har styr på inden vi begynder dags fordi jeg skal prøve sådan producere de der rapporter for eksempel Daniel sagde okay nu kommer vi til at arbejde med Claus det er nu afdeling for mig så ikke vi vil jo bare have et nu Vi vil kigger på hvordan dette ser ud som Daniel har også vist om det er faktisk den der struktur om vi kæter og vi begynder faktisk lave rapporter eller er vi nødt til at oplære dem hvordan man opsamle det eller måske hjælper dem med nogle programmer for om det kan måske lave måske kan vi jo sådan lave farve app så for at gøre det automatisk."

"Daniel sagde i hvert fald ikke må og så nu hvis jeg skal vurdere så nu er vi meget på at få strukturer få det i rapporter så mennesker kan se det på et højt overblik det er der hvor vi nu fremtiden når alle sammen har rapporter som de har brug for så det er nok lidelsen som vil bestemme hvad næste det ikke skal stoppe."

Dataarbejde

“så er det blevet lidt mere systematiseret her”

“Er der en grund til at de siger køb per tilskuer eller omsætning på tilskuer og ikke omsætning for køb?”

“Er det den kan ja det ved jeg simpelth ikke men jeg mener det er 100 og nogle kroner eller sådan noget der er i i kassen der. Men grunden til den er også så svær og nok grunden til vi ikke arbejder så meget på den”

“Nej, men altså jeg jeg arbejder i Excel, men altså det Olga arbejder 100 gangs meget.”

“Altså for eksempel partnerbilletter, nu kan jeg simpelthen ikke huske det, men jeg tror det er, tror det er 1800, når man regner ud på snit, at der er 1800 billetter er egentlig solgt til hver eneste kamp, så så jeg skulle helst have plads til dem.”

“hvor er jeres ø målsætning komme hen med alle de her nye systemer? Øh, systemer de mener PowerBI og Ja, PowerBI og alt detta benytter. Så lige nu er det mange manuelle processer,”

Kapacitetsudnyttelse

“Det skal helst ligge på omkring 75% af de tilskuere, der er. øh de skal handle. Så det det er mere det tal der vi bruge”

“målsætning skal laves fire uger før eller sådan noget.”

“vi i høj grad gør, fordi altså lad os nu bare sige, der er en virksomhed som har 10 billetter øh for eksempel til hospitality til hver eneste kamp på det tidspunkt de har lavet aftalen”

“6200 øh tilskuere til typisk på kampene pt. Hvor hvor tæt pakket skal stadion være før den før det før det er en fed oplevelse? Jeg tænker, der må være en eller anden kritisk masse for, hvornår det ser dødt og kedeligt ud, fordi der er for få folk. Hvornår rammer man den grænse? Altså at lige nu ligger vi jo så op sjovt nok på de der 62%, ikke?”

“Altså jeg vil sige sådan, jeg tror jeg tror rammen omkring 5000 Så så er det et ret fedt stand.”

Billetaktivering og fremmøde

“Ja men det er altså hvad skal man sige vores billetdat altså vores billetdat platform der det er altså det er et univers, hvor der er en masse statistikker og sådan nogle ting til Kan godt bare arbejde i billetsystemet”

“Altså, det er jo ikke fordi vi sådan hvis du tager det her år, det er jo ikke sådan fordi vi har blæst hele ligaen væk, men altså vi ligger alligevel stadigvæk med 6200 og hvor det var 6400 sidste år.”

“Altså hvis vi er, hvis vi er hvis vi er 8500 her, jamen så så er det måske maks 3.000 eller to, et5 eller sådan noget, som fysisk har købt billetten til kampen. De øvrige, de har købt billetten via partnerskaber eller via abonnementer. Så det er jo ikke nødvendigvis bare lige til den ene kamp der. Det er jo til en serie af kampe.”

“det er at vi forsøger som jeg sagde på abonnementerne og det Der forsøger vi at sælge menuer med ind.”

“det kører på kamp det kører på kamp på hvad skal man sige tilskuerfordeling på kampen så ved jeg at for eksempel sidste gang til afland der var der 625 billetter afhentet og 462 blev benyttet”

“Vi får lige penge i kassen, der skal. Problemet er jo bare, hvis vi gang på gang kan se de her 10 billetter, de ikke bliver udnyttet, så får de jo ikke det udbytte ud af netværket, som de skal have.”

“men det er jo altså for eksempel af aktiveringen, hvor hvor du ikke betaler noget for billetten. Men det det ved du selv. Hvis du hvis du har adgang til at hente en billet, jamen så å jeg tror fandme også min lillesøster og hendes kæreste og deres to børn, de vil nok også god med fordi du har adgang. Det koster da ikke noget, så booker du otte billetter i stedet for fire. Men din lillesøster, hun gider sgu ikke med alligevel. Så så bliver fremmødet jo mindre. Og det er derfor sådan noget aktivering af det der, vi snakker ekstremt meget om, om vi bare skal putte et lille lille beløb på. Altså det kan være, at al aktivering i fremtiden godt kunne være, at når du har hentet billetten, så betaler du den, så får du en fransk hotdog eller en sodavand eller et eller andet, således at man stadigvæk i princippet betaler nul kroner, men du betaler noget for billetten, fordi Har vi bare haft på en art, så har vi altså en helt anden motivation til at bruge den”

Økonomi

“Altså jeg tror der hvor vores udsving er størst i forhold til tid og vejr, det tror jeg det er, det er bås øhm vi vi vores volumen, vores økonomi er er markant. større end en god hvad er det det hedder sommerdag efterårsdag altså eller undskyld sommerdag måske fredag aften og sådan nogle ting er markant større altså jeg tror da vi da vi har fck øh og sønder fredag aften her i starten der mener jeg vi ligger på en 76,8 kr. i omsætning per tilskuer hvor at vi mod hvad er det det hedder OB nu her er helt nede på 47 så der der er stor forskel men det er også meget logisk ikke fordi Altså køber du en fadøl kontra en kop kaffe, så er der i hvert fald en 20årig i forskel, ikke? Så noget er meget”

“Øh, og til en normal kamp, der har vi på vores partnerside og på vores menuside i abonnement, der har vi solgt omkring 1000 menuer. Og når de bliver købt i båen, så er det jo nulkroners omsætning. Så derfor bruger vi meget mere. Vi bruger omsæ, altså salget per tilskuer, og så bruger vi antallet af handler. Og antallet af handler, antallet transaktioner.”

“vi har udsolgt på men jo mere du betaler, jo mere værdi der er på dit produkt, desto højere er dit fremmøde. Og så derfor dem der betaler, dem der har bedst fremmøde, det er dem der har købt til hospitality. Det er omkring hvad det er 1250 kr. hver billetkost om ikke? Så så dem er der højest fremmed på. Og så er det de sæsonkortholdere, som sidder midt på ovre i øst i de på de på de dyrere pladser. Men altså der er forskel på, om du betaler 39 kr. for en barnebillet på syd, og så du betaler, hvad er det hedder 189 for en voksen billet. med menu ovre på øst, så bruger du altså den der til 189 mere.”

Systemlandskab

“Den krølle der får vi faktisk på hvad er det det hedder på søndag kommer der en ny øh en ny en der gerne vil leverere kassesystemer til os som ø som hvad skal man sige skal prøve at sætte de sætter de lav to test terminaler til den her gang og fire til de næste. Og det er meget fordi vi ikke kan få nok data ud af vores nuværende øh at vi at vi ønsker at skifte ikke Ja.”

“vi alt vores aktivering sker sker via vekkode. Så alle der hvad skal man sige bruger de her aktiveringer, de

er opretter en profil, så vi ved hvem vi snakker med og så bruger de vekkoderne. Så vi ved altid via koderne hvordan tilgangen den er til billet."

"Så er det spørgsmål, hvor mange der trækker billetterne. Men hvis man kigger på vores partners sideide, så så deres adgang til billetter, det op i to kategorier. Der er der er det, man kalder eventbilletter. Det vil sige, at de har to billetter til alle kampe. Altså to billetter til hver kamp. Så har de også noget, der hedder puljebilletter. Der kan de for eksempel have 20 puljebilletter, men de bestemmer selv, hvornår de bruger de puljebilletter. Så for eksempel i en A-kamp, så kan de vælge at bruge, jeg tror, vi har en grænse på seks lige indtil videre, medmindre man lige spørger, om man må bruge 10. Altså, men så så derfor så det det er egentlig den måde, at der kan være forskellen på, men det er mere, hvordan virksomhederne selv vælger at bruge deres puljebilletter. Altså eventbilletterne, det er to til hver eneste kamp for eksempel. Så en del af rammen er den samme."

"Det det eneste de eneste der har der har fast, hvad skal man sige, pladser altså som som har en en en et sæde på stadion. Det er det er dem der har årssæsonkort og hvad det sæsonkort abonnementer. Og så er det øh 250 partnere som har faste pladser ude i de sorte sæder. Ellers alt andet er billetttræk. Altså det vil sige du går ind og aktiverer. Du kan godt have fire pladser til kampen du kan bruge. Og hvis du ikke går ind og aktiverer dine pladser, så er pladserne ikke optaget. Så bliver de ikke taget."

Placeholder

"så det ville v have en betydning på de billetter, som ikke er solgt i forvejen, at hvis der sker noget området."

"Mere om hvis der er et eller anden stor begivenhed, der gør at folk de måske fravælger, Smukfest for eksempel."

"Altså der er mange steder, hvor vi vores kommunikation, og hvor vi vores opfordring til at frigive pladser, og hvor vi vores opfordring til at udnytte de pladser, de har. Når vores fokus bliver fjernet en lille smule, for det er måske mere over på end det nye appunivers, vi skal have fortalt om. Og Øh, vi laver meget mere content i forhold til aktivering af videomæssig og sådan nogle ting."

"Jamen vi starter egentlig med at kigge på kapaciteten, altså vi har på stadion, og så starter vi med at og så tager vi og forholder de, hvad er det det hedder den struktur vi laver, fordi vi har vi har Nogle opdelinger. Altså vi har, hvad er det det hedder en sydtribune, stemmeskabetribune, der er nogle begrænsninger på debillettyper de har. Vi har en østtribune ovre på den side derovre, som er meget til sæsonkortholder. Der er nogle begrænsninger på den. Så har vi noget partnerafsnitter. Så det er sådan meget, hvad der hvad der er solgt ind til de enkelte områder"

"hvad skal man sige skabe fællesskab omkring det at opleve tilskuertal, fordi vores tilskuertal det er mega vigtigt. Altså fodbold det er for fællesskaber. Fællesskaber det er sjovest når vi er mange. Og det vil vi rigtig rigtig gerne. Så der vi har så mange afledte parametre af, at det lykkedes os at have et højt tilskuertal. Derfor er det vigtigt, at alle afdelinger på en eller anden måde har en fornemmelse af, de bidrager. Og så kan du sige kommunikation og sådan nogle ting. Jamen, hvordan hvordan er de lige mål, at de at de bidrager? Og det har vi bare sådan helt målart forsøgt at sige, jamen for eksempel, hvad er det det 720 i BD, så må vi jo kigge på, hvad er det, hvorfor

opnår vi det ikke? Jamen, Øh, har vi ikke solgt pladserne? Jo, det havde vi faktisk. Altså, vi kunne godt have noget 720, fordi pladserne var med,"

"Så der tag der der skal vi ud og have fat i virksomhed og snakke, hvad er hvad er det altså At er det ikke fedt nok det vi har, eller har vi en forkert approach på? Skal vi have flyttet i skybene? Skal vi have lavet om, at I har to skybaser hver eneste gang eller sådan noget? Så det skal altså det skal vi ud og have, fordi og vi skal gøre det før at vi sidder ude til kundemødet og siger, prøv at hør, vi forlænger sgu ikke sponsoratet, fordi at vi bruger det ikke. Der skal vi have fanget det. Jeg vil egentlig gerne dele den og sige på et dashboard der der skal det sådan være meget, hvad der er relevant for kontoret."

"Ja. Øhm. Og det vil sige, det er vores det er meget op på vores kopimål, hvor vi er henne i forhold til vores kopimål, som er vigtigt, som er det over Altså jeg tror vi havde en snak Olga i i går med med hvad er det det hedder i forhold til clubizer ikke hvor hvor vi sådan lidt ligesom siger altså Olga har lavet rigtig rigtig mange opsummeringer på forskellige produkter der er herude på stadion"

"Altså der er jo noget med det er naboens dreng, som man ved hvem er, som har løbet rundt og spillet fodbold med ens egen søn. Altså det er der noget over. Så er der noget over de der meget spændende dygtige spillere, hvor Thomas er et godt eksempel, og så er der nogen der bare godt kan lide at se nogen, der kan noget flæ og drible og sådan noget"

"men FCK de står jo i samme situation, men måske altså man bare et meget højere højt tilskrtal, så derfor er der også flere der efterspørger nogle forskellige. Men de gik simpelthen ud og sagde, prøv at høre her, vi kan vi vi har nogle kerneprodukter, vi kan lave, så vi kan vi kan lave til masserne øh selvfølgelig også med andre produkttyper, men til masserne, dem der har helt specielle ønsker, som vi egentlig også skal det, det kan vi ikke gøre på stadion. Så vi laver en aftale med volt. Så kan man bestille ud i byen, og så er der et sted, hvor man kan få leveret det. Og ved du hvad? Alle er glade. Og det har vi betydet, at de har fået så et hjørne, hvor der er nogen der bestiller champagne og caviar hver gang. Det tror jeg ikke, vi kommer til at have. Men men jeg tror, det er sådan noget, vi kommer til at kigge ind på, for vi kan ikke vi kan ikke gøre det andet."

D.6 Interview med Daniel og praktikanterne

Datakultur

"Vi tager en afdeling, og så begynder vi at arbejde med det. Altså selvfølgelig hvis de har brug for nogen data og de har nogle andre ønsker, så de vil ikke gå til ind fra marketing og så siger, at nu har vi lyst til det her, så de går til Daniel eller til mig eller vi kan godt lide at det er faktisk bare struktur, fordi det er meget nemmere at arbejde, fordi som den Daniels er vi også meget Vi er presset med tiden, og hvis vi har sådan mere tid, så vi har også sådan flere opgaver, som vi synes faktisk, at det ville være ikke spændende, men nødvendigt måske til at bare for at å få et luft op vores det måhed, men så det vil sige, at vi kan godt lide noget det struktureret, og hvis de har lyst til at snakke om noget, så vi meget sådan vi opføer dem, at de booker et møde hvor vi hedder ikke underviser, men forklare dem, hvordan det fungerer. Hvad er vores krav? Det er meget øh Daniels PowerPoint

Daniel plejer også til at vise dem og så forklarer den der og så vi også sådan fraråde dem at hvis de skifter deres mening næste dag så måske er det ikke en god ide fordi vi er nødt til at starte forfra"

"Ja, der indkalder vi den afdeling, vi samarbejder med, fordi at for eksempel så kan de sige, vi vil gerne have det her dashboard, det skal se sådan her ud. Og så kigger vi i data og siger vi, det her data kommer aldrig nogensinde til at kunne det. Altså som i aldrig nogensinde. Det kræver, vi gør lige nøjagtigt de her ting her. Så det har vi valgt at implementere, det kommer til at påvirke jeres arbejde lige nøjagtigt sådan her. Så I skal gøre sådan her, sådan her og sådan. Og så laver vi en guide til dem. Og så det er sådan her det skal gøres frem."

"Tror det startede meget som at h man kigger lidt på det sådan okay, jeg prøvede at fremlægge sådan lidt nogle problemstillinger i forhold til medarbejderafhængigheder og virksomhedsafhængighed. Og så byggede vi ligesom så startede vi i en enkelt afdeling og sagde, okay, hvad kan vi gøre her? Men så i takt med at, hvad kan man sige, slubbrugeren sådan en som Palle begynder at se, hvad er det faktisk man kan"

"så kommer der et efterspørgspråmere og mere og mere, og så spreder det sig. I takt med at så har Palle lige pludselig et værktøj, som nogen i marketing synes er fedt, jamen så kunne de også godt tænke sig noget, der minder om. Så så helt til at starte med var det nok meget boret af dataafdeling, og nu er det meget mere. Vi skal faktisk prøve at sørge for, at vi bare kan nå at holde fast i eller Nå de projekter vi har, fordi vi har rigelige opgaver, fordi det spreder sig, fordi man oplever, okay, det jeg har brugt fire timer før hver eneste dag på, det er for ham her løst på to minutter. Det vil man jo gerne. Og oven i købet ser det pænere ud."

Dataarbejde

"hvad er det for nogle udfordringer vi egentlig har i virksomhed og hvordan skal de løses og hvorfor er det vi ikke har løst dem i dag altså sådan meget visuelt forvist jamen okay Vi har de her processer. Det tager faktisk fire timer for en medarbejder at gøre hver dag. Vi har tre af de her medarbejdere. Det er 12 timer om dagen, som bliver brugt på noget, som er fuldstændig manuelt. Det skal fungere automatisk."

Data governance og struktur

"Jeg tror det man kan sige er, hvis dataafdelinger er involveret i projektet, så er det også der står for. Men vi har jo også projekter slagsafdelinger, vi ikke kan understøtte i dag, fordi vi bemanding til. Øh, men grundlæggende så hvis vi er inde over et projekt for eksempel med Palle, så så tager vi det ansvar øh og hvis vi ikke kan have indtastningsarbejdet af whatever reason, det kan der jo godt være nogle gange, at vi ikke sidder som den manuelle til at taste ting ind, så sørger vi for at dem der sidder med indtastningsarbejde får at vide nok arbejde, så vi sikrer os at det bliver gjort på den måde som vi aftaler."

Placeholder

"Øh vi mangler nogen der har de kompetencesæt som dem der sidder herovre og har. Okay."

"Øh i forhold til det er nogen der skal kunne arbejde med datastruktur for rensede data. De skal have nogle kodekompetencer og så skal man selvfølgelig også have nogle menneskelige kompetencer fordi man kommer til at sidde meget tæt på stakeholder."

"Det det er sådan på kompetence nu. Hvis vi er den organisation, vi er lige nu fuldstændig statisk også om to år, så der er ikke blevet ansat nogen, der er ikke blevet lavet nye afdelinger, så tror jeg på at øh fem fuldtidsansættelser kunne understøtte det ud fra de behov."

"Men hvad hvad med den administrative del hvordan?"

"Altså nu har jeg siddet i med nogle data, hvor jeg primært har renset data og stillet nogle spørgsmål løbende, og jeg har ikke talt så meget, kan man sige, direkte datamodenhed med den afdeling. Øhm, men det virker til, at de er klar på forandringen, og de gerne vil det. Øhm, og om jeg jeg jeg måske heller ikke den rette person at spørge i forhold til sådan at rate det fra et til 10. Men jeg jeg tror også at øh jeg vil give dig ret i at det er det der med de øh de er meget åbne over for data, men jeg tror også det er en blanding af at finde ud af hvad er det data kan at det er sådan lidt lidt der mange af afdelingerne er og så hvor hvordan man arbejder med det nu."

"Jeg tror jeg skal have specificeret dig hele til 10er, fordi hvis vi siger, at 10 det er den bedste i Superliga, så vil jeg måske sige, så er vi måske syv eller otte sådan realistisk."

"Og hvad med for eksempel markedsføring afdeling og hvordan vi altså marketing, altså den der hedder mark, den der hedder salg, de refererer til hvad hedder det vores direktør dem der refererer til økonomidirektøren det er hvad hedder det os Palle øh HR og økonomi. Vi refererer til økonomidirektøren og resten refererer til den administrerende direktør og så er der nogle mellemledere kan man sige mellem"

E Bilag 5: Dataklargøring

E.1 Indlæsning af pakker

Først indlæses de nødvendige R-pakker til databehandling, web scraping, database håndtering og moving averages.

```
# Load packages
pacman::p_load(tidyverse, rvest, janitor, RSQLite, slider)
```

E.2 Web scraping af Superliga data

Data for alle Superliga-sæsoner fra 2002-2025 hentes fra superstats.dk ved hjælp af web scraping. Dataene indeholder kampoplysninger som dato, resultat, tilskuertal og dommere.

```
# superstats data ------------------------------------------------------


# link til superstats, uden sæson år (kommer i loop funktion)

url <- "https://superstats.dk/program?season="


# laver en tom liste som tabellerne skal ligge i
```

```r
seasons <- list()

# loop hvor hver sæson kommer som en liste, med runder som tabeller
# alle listerne sættes i den tomme liste

for (i in 2002:2026) {
  html <- read_html(paste0(url, i), encoding = "UTF-8")

  tables <- html |>
    html_element("#content") |>
    html_elements("table") |>
    html_table()

  seasons[[as.character(i)]] <- tables
}

# renser navne på tabellerne med janitor for at kunne arbejde videre med dem

for(season_år in names(seasons)) {
  seasons[[season_år]] <- lapply(seasons[[season_år]], clean_names)
}

# laver to ny variabler. en for sæson år og en for runde nummer

for(season_år in names(seasons)) {
  seasons[[season_år]] <- lapply(seasons[[season_år]],
                                 \(df) mutate(df, season_year = season_år,
                                              round = parse_number(names(df[1]))))
}

# navngiver tabeller med ordentlige variabelnavne, sådan at de kan samles i
# en enkelt dataframe

for(season_år in names(seasons)) {
  seasons[[season_år]] <- lapply(seasons[[season_år]],
    \(df) df |>
      setNames(c("ugedag", "dato_tid", "kamp", "resultat", "tilskuere",
```

```
                "dommer", "tv", "sæson_år", "runde_nr"))
  )
}

# samler tabellerne i en enkelt dataframe

seasons_all <- seasons |>
  purrr::flatten() |>
  bind_rows()

# får fejl da det ikke er alle observationer i ugedag og kamp der er samme type
# alle observationer i ugedag og kamp tvinges derfor til at være character
# kør ovenstående igen

for(season_år in names(seasons)) {
  seasons[[season_år]] <- lapply(seasons[[season_år]],
    \(df) df |> mutate(ugedag = as.character(ugedag),
                       kamp = as.character(kamp))
      )
}

# samler tabellerne i en enkelt dataframe

seasons_all <- seasons |>
  flatten() |>
  bind_rows()

# tilskuere blive i ovenstående funktion lavet til en double, men står ikke længere i 1000'er
# tilskuere variabel laves derfor i 1000'er i stedet for

seasons_all <- seasons_all |>
  mutate(tilskuere = tilskuere * 1000)
```

E.3 Upload til SQLite database

De indsamlede data gemmes i en SQLite database for nem adgang og vedligeholdelse.

```
# Uploader alle superstatsdata til SQLite database
con <- dbConnect(SQLite(), "data/fodbolddata.sqlite")
dbWriteTable(con, "db_seasons_all", seasons_all, overwrite = TRUE)

# tjekker om den er blevet uploadet
dbListTables(con)

dbDisconnect(con)
```

E.4 Oprettelse af variabler

Nye variabler oprettes baseret på de eksisterende data, herunder VFF's mål, resultater, point og moving averages for de seneste kampe.

```
# først laves en række nye variabler ud fra de fulde datasæt for superligaen, sådan at
# variabler som f.eks antal mål i sidste kamp, blive den reele seneste kamp, og ikke seneste hjemmekamp
# og så der kan laves totale aggregerede point for hver sæson

# henter data alle superliga kampe, ikke kun vff hjemmekampe, fra database

con <- dbConnect(SQLite(), "data/fodbolddata.sqlite")

seasons_all <- dbReadTable(con, "db_seasons_all")

dbDisconnect(con)

# finder alle VFF's kampe, både hjemme og ude

seasons_all <- seasons_all |>
  filter(str_detect(kamp, "VFF-") | str_detect(kamp, "-VFF"))

# opdeler resultat variablen i to, sådan at hjemmeholdets mål og udeholdets mål
# står for sig selv

seasons_all <- seasons_all |>
  separate_wider_delim(
```

```
    resultat,
    delim = "-",
    names = c("hjemme_mål", "ude_mål"),
    too_few = "debug"
  )

seasons_all <- seasons_all |>
  mutate(
    vff_mål = as.numeric(if_else(str_detect(kamp, "VFF-"), hjemme_mål, ude_mål)),
    .after = kamp
  ) |>
  mutate(
    mål_sidste_kamp = as.numeric(lag(vff_mål)),
    .after = vff_mål
  )

# ny variabel for mål i de sidste tre kampe, ved at bruge slider package, og lagged version

seasons_all <- seasons_all |>
  mutate(
    mål_sidste_tre = slide_dbl(vff_mål, sum, .before = 2, .complete = TRUE)
  )

seasons_all <- seasons_all |>
  mutate(
    mål_sidste3_lagged = lag(mål_sidste_tre)
  )

# laver resultat og lagged resultat variabel

seasons_all <- seasons_all |>
  mutate(
    vff_resultat = if_else(str_detect(kamp, "-VFF") & hjemme_mål > ude_mål, "tabt", NA),
    .after = kamp,
    vff_resultat = if_else(str_detect(kamp, "-VFF") & hjemme_mål < ude_mål, "vundet", vff_resultat),
    vff_resultat = if_else(str_detect(kamp, "-VFF") & hjemme_mål == ude_mål, "uafgjort", vff_resultat),
    vff_resultat = if_else(str_detect(kamp, "VFF-") & hjemme_mål > ude_mål, "vundet", vff_resultat),
```

```
    vff_resultat = if_else(str_detect(kamp, "VFF-") & hjemme_mål < ude_mål, "tabt", vff_resultat),
    vff_resultat = if_else(str_detect(kamp, "VFF-") & hjemme_mål == ude_mål, "uafgjort", vff_resultat),
    vff_resultat_lagged = lag(vff_resultat)
  )

# laver ny variabel for antal point, der skal bruges til at lave variabel for
# point i de sidste tre kampe, og aggregered point for sæsonen

seasons_all <- seasons_all |>
  mutate(
    point_kamp = if_else(str_detect(vff_resultat, "tabt"), 0, NA ),
    .after = vff_resultat_lagged,
    point_kamp = if_else(str_detect(vff_resultat, "uafgjort"), 1, point_kamp),
    point_kamp = if_else(str_detect(vff_resultat, "vundet"), 3, point_kamp)
  )

# ny variabel for antal point i de sidste tre kampe, ved at bruge slider package, og lagged version

seasons_all <- seasons_all |>
  mutate(
    point_sidste3 = slide_dbl(point_kamp, sum, .before = 2, .complete = TRUE),
    .after = point_kamp
  )

seasons_all <- seasons_all |>
  mutate(
    point_sidste3_lagged = lag(point_sidste3)
  )

# ny variabel for aggregerede point for hver sæson, og lagged version
# igen bruges slider package

seasons_all <- seasons_all |>
  group_by(sæson_år) |>
  mutate(
    point_sæson = slide_dbl(point_kamp, sum, .before = Inf, .complete = FALSE)
  ) |>
```

```
  ungroup()

seasons_all <- seasons_all |>
  mutate(
    point_sæson_lagged = lag(point_sæson)
  )
```

E.5 Håndtering af dato og tidspunkt

Dato og tidspunkt konverteres til standardiserede formater, og der oprettes variabler for hvilket tidspunkt på dagen kampen spilles.

```
# datotids variabler
# inden den samlede datotid variabel laves, skal det reele år kampene blev spillet først laves,
# i stedet for sæsonåret. derefter laves tidsvariablen, som også konverteres til en ny variabel
# hvor tidszonen er UTC, som skal bruges til at få data fra DMI

seasons_all
 mutate(
    real_år = if_else(str_detect(dato_tid, "/(07|08|09|10|11|12) "),
                      as.numeric(sæson_år) - 1, as.numeric(sæson_år))
  ) |>
  mutate(datotid = ydm_hm(paste(real_år, dato_tid), tz = "Europe/Copenhagen")) |>
  mutate(datotid_utc = with_tz(datotid, tzone = "UTC"))

# Laver ny variabel for hvilket tidspunkt på dagen det er
seasons_all
  mutate(
    tidspunkt = if_else(hour(datotid) %in% 10:14, "middag", NA),
    tidspunkt = if_else(hour(datotid) %in% 15:17, "eftermiddag", tidspunkt),
    tidspunkt = if_else(hour(datotid) %in% 18:23, "aften", tidspunkt)
  )
```

E.6 Historiske møder mellem hold

Der oprettes variabler der indeholder information om sidste møde mellem de to hold, herunder tilskuertal og resultat.

```
# nye variabler for sidste møde mellem holdene. ved at bruge group_by på kamp variablen,
# bliver det pr sidste hjemmemøde og sidste udemøde. altså bliver sidste møde tilskuere altså en
# variabel for hjemmekampene hvor det er tilskuere i sidste hjemmekamp, og omvendt for udekampe
seasons_all
  group_by(kamp) |>
  mutate(
    sidste_møde_tilskuere = lag(tilskuere)
  ) |>
  ungroup()

# den nye variabel for resultatet for det sidste møde skal ikke opdeles pr. sidste hjemmemøde eller ude
# denne skal bare være generelt for sidste møde. derfor findes først modstanderen, og grupperes efter d
# og derefter laves variablen for resultatet i det sidste møde
seasons_all
  mutate(kamp2 = kamp) |>
  separate_wider_delim(
    kamp2,
    delim = "-",
    names = c("hjemme", "ude")
  ) |>
  mutate(
    modstander = if_else(str_detect(hjemme, "VFF"), ude, hjemme)
  ) |>
  group_by(modstander) |>
  mutate(sidste_møde_resultat = lag(vff_resultat)) |>
  ungroup()

# da der ikke findes alt relevant vejr data før 2003, fjernes den tidligere sæson
seasons_all
  filter(as.numeric(sæson_år) >= 2003)
```

E.7 Formatering til DMI API

Dato og tidspunkt formateres til DMI API's specifikke format, så vejrdata kan hentes for hver kamp.

```
# for at lave et loop der kan hente DMI data for alle kampdagene på en gang,
# skal en ny variabel laves ud fra datotid_utc variablen
# den splittes sådan at selve datoen står for sig selv og selve klokken for sig selv
# de samles igen i en ny variabel med et T efter datoen, og et Z efter klokken,
# for at få det i det format der skal bruges i DMI's API request.
# med dette kan der laves et loop.

seasons_all
  mutate(datotidutc = datotid_utc) |>
  separate_wider_delim(
    datotidutc,
    delim = " ",
    names = c("dato_dag_only", "dato_tid_only")
  ) |>
  mutate(dmi_dato = paste0(dato_dag_only, "T", dato_tid_only, "Z"))

# datasættet med de nye variabler uploades i databasen

con <- dbConnect(SQLite(), "data/fodbolddata.sqlite")

dbWriteTable(con, "db_seasons_all", seasons_all, overwrite = TRUE)

dbDisconnect(con)
```

E.8 DMI data

Vejrdata hentes fra DMI's API for alle VFF hjemmekampe, baseret på kampenes tidspunkt.

```
# dmi data ----------------------------------------------------------------
# for at lave loopet vælges først kun vff hjemmekampe, og kampe der ikke har et resultat
# og ikke er relevante fjernes. med dette og datotids variable i det nye format, kan der laves et loop

vff_hjemme <- seasons_all |>
  filter(str_detect(kamp, "VFF-")) |>
  filter(resultat != is.na(resultat)) |>
  filter(resultat != "Optakt" )

# vælger den nye dmi_dato og trækker den ud som en vektor, som skal bruges til at få data fra DMI

dmi_dato <- vff_hjemme |>
  select(dmi_dato) |>
  pull()

# ved at bruge disse tidspunkter til at trække data ud fra dmi, fåes der ikke komplet data
# ved halve timer fåes der færre data punkter og variabler ud end for de hele timer,
# og for de få kampe der er startet xx:35, eller xx:05 fåes der slet ikke noget data
# derfor laves alle tidspunkter om til hele timer

# ændrer alle tider til nærmeste hele time, ved at runde ned gennem floor_date,
# og laver igen datoen til DMI API formatet

seasons_all <- seasons_all |>
  mutate(datotid_utc = floor_date(datotid_utc, unit = "hour")) |>
  select(-dato_dag_only, -dato_tid_only) |>
  separate_wider_delim(
    datotid_utc,
    delim = " ",
    names = c("dato_dag_only", "dato_tid_only")
  ) |>
  mutate(dmi_dato = paste0(dato_dag_only, "T", dato_tid_only, "Z"))

# datasættet uploades igen i databasen, for at kunne bruge den nye dmi_dato til at joine på
```

```r
con <- dbConnect(SQLite(), "data/fodbolddata.sqlite")

dbWriteTable(con, "db_seasons_all", seasons_all, overwrite = TRUE)

dbDisconnect(con)
```

E.9 Processering af vejrdata

Vejrdata fra DMI API'et ekstraheres og konverteres til et struktureret dataframe format.

```r
# dmi datoen trækkes ud som en vektor, sådan at den kan bruges i et loop

vff_hjemme <- seasons_all |>
  filter(str_detect(kamp, "VFF-")) |>
  filter(resultat != is.na(resultat)) |>
  filter(resultat != "Optakt" )

dmi_dato <- vff_hjemme |>
  select(dmi_dato) |>
  pull()

# vi har nu alle kampdagene, i et format der kan bruges i et loop
# bygger url op. datoen i request url kommer i for loopet

base_url <- "https://dmigw.govcloud.dk/v2/"
info_url <- "metObs/collections/observation/items?"
req_url <- "stationId=06060&datetime="
limit <- "&limit=100000"
api_key <- Sys.getenv("MY_API_KEY")

# laver en tom liste som data skal ligge i

dato_vejr <- list()

# loop hvor vejr data for alle kamptiderne hentes

for (dato in dmi_dato) {
 full_url <- paste0(base_url, info_url, req_url, dato, limit, "&api-key=", api_key)
```

```
 api_call <- httr::GET(full_url)

 api_JSON <- httr::content(api_call, as = "parsed", simplifyVector = TRUE)

 dato_vejr[[as.character(dato)]] <- api_JSON
}

# vejrdataen ligger i dataframes, som ligger i andre dataframes, som ligger i lister,
# der ligger i en liste. der er heldigvis mønstre og ens navngivninger, så
# der laves nemt en funktion der gennem lapply køres på alle listerne

properties <- lapply(dato_vejr, function(x) x$features$properties)

# samler alt vejr data i en enkelt dataframe i stedet for en liste

vejr_alle <- bind_rows(properties)

# lægger alle vejrobservationer i SQLite database, inden der udvælges første variabler,
# i tilfælde af at andre/flere variabler skal bruges

con <- dbConnect(SQLite(), "data/fodbolddata.sqlite")

dbWriteTable(con, "db_vejr_alle", vejr_alle, overwrite = TRUE)

vejr_alle <- dbReadTable(con, "db_vejr_alle")

dbDisconnect(con)
```

E.10 Udvælgelse af relevante vejrvariabler

Fra det komplette vejrdatasæt udvælges de mest relevante variabler: nedbør, temperatur og vindhastighed.

```
# udvælger vejr variabler og lægger i en ny dataframe

vejr_udvalgt <- vejr_alle |>
  filter(parameterId %in% c("precip_past1h", "temp_dry", "wind_speed")) |>
  select(parameterId, value, observed)

# pivot sådan at det står rigtigt som variabler i stedet for observationer

vejr_udvalgt <- vejr_udvalgt |>
  pivot_wider(
    names_from = parameterId,
    values_from = value
  )

# lægger de udvalgte vejrobservationer i SQLite database

con <- dbConnect(SQLite(), "data/fodbolddata.sqlite")

dbWriteTable(con, "db_vejr_udvalgte", vejr_udvalgt, overwrite = TRUE)

dbDisconnect(con)
```

E.11 Helligdage data

Data om danske helligdage hentes fra date.nager.at API'et for at identificere kampe på helligdage.

```
# date.nager data ----------------------------------------------------------

# bygger url op. request url'en består kun af årstaller, og kommer i loopet

base_url_dn <- "https://date.nager.at/api/v3/PublicHolidays/"
country_code <- "/DK"

# lave en tom liste som data skal ligge i

holidays <- list()

# loop hvor alle helligdagene i årene 2002 - 2025 hentes

for (i in 2002:2025) {
 full_url <- paste0(base_url_dn, i, country_code)

 api_call <- httr::GET(full_url)

 api_JSON <- httr::content(api_call, as = "parsed", simplifyVector = TRUE)

 holidays[[as.character(i)]] <- api_JSON
}

# samler data i en enkelt dataframe i stedet for en liste, omdøber variabler
# og vælger kun datoen, og hvilken helligdag det er

holidays <- holidays |>
  bind_rows() |>
  mutate(
    dato = date,
    helligdag = localName
  ) |>
  select(dato, helligdag)
```

```
# uploader helligdage data til database

con <- dbConnect(SQLite(), "data/fodbolddata.sqlite")

dbWriteTable(con, "db_holidays", holidays)

dbDisconnect(con)
```

E.12 SQL query

Alle datasæt (kampe, vejr, helligdage og billetsalgsdata) sammenkobles til ét komplet analysedatasæt.

```
# joins med andet vff data ------------------------------------------------

con <- dbConnect(SQLite(), "data/fodbolddata.sqlite")

# undersøger først hvad der ligger i db_vff for at se hvordan og hvad der skal joines

db_vff <- dbReadTable(con, "db_vff")
dbListTables(con)

# det er kun d10-, d7- og d3- tilskuere der mangler i det superstats datasættet
# derfor vælges kun disse og joines med en compound key bestående af år og runde
# samtidig joines det med dmi og date.nager data

vff_all <- dbGetQuery(con,
  "SELECT s.*, r.temp_dry, r.wind_speed, r.precip_past1h, g.helligdag,
   t.d10_tilskuere, t.d7_tilskuere, t.d3_tilskuere
   FROM db_seasons_all AS s
   LEFT JOIN db_vejr_udvalgte AS r
   ON s.dmi_dato = r.observed
   LEFT JOIN db_holidays AS g
   ON s.dato_dag_only = g.dato
   LEFT JOIN db_vff AS t
   ON s.real_år = t.år
   AND s.runde_nr = t.runde
   WHERE s.kamp LIKE '%VFF-%'"
)
```

```
dbDisconnect(con)
```

E.13 Endelig datarensning

Det sammenkoblede datasæt renses for manglende værdier, og der oprettes nye variabler som sommerferie, måned og uge.

```
# nu når alle datasættene er joined og hentet fra databasen rydder vi lidt mere op i det
# de kampe der ikke har et resultat og ikke er relevante fjernes, og irrellevante variabler fravælges

vff_all <- vff_all |>
  filter(resultat != is.na(resultat)) |>
  filter(resultat != "Optakt" ) |>
  mutate(datotid = as.POSIXct(datotid)) |>
  select(-dato_tid, -kamp, -vff_resultat, -point_kamp, -hjemme_mål, -(resultat:resultat_remainder), -do
         -mål_sidste_tre, -point_sæson, -dato_dag_only, -dato_tid_only, -dmi_dato, -real_år,
         -point_sidste3, -vff_mål, -ude_mål, -hjemme, -ude)

# der er stadig NA'er i de to sidste_møde variabler, så disse skal fikses
# for at fikse sidste_møde_tilskuere, laves først en ny variable der kategoriserer efter om
# sidste_møde_tilskuere er NA eller ej

vff_all <- vff_all |>
  mutate(nyt_hold = if_else(is.na(sidste_møde_tilskuere), 1, 0))

#filtrerer efter disse kampe, for at undersøge dem nærmere

nye_hold <- vff_all |>
  filter(nyt_hold == 1)

# der er tydelig forskelle på størrelsen af holdene, så de deles op sådan at dem havde mindre end 4000
# tilskuere får egen kategori, og dem havde mere end 4000 tilskuere får egen kategori

vff_all <- vff_all |>
  mutate(lille_stor_ny = case_when(nyt_hold == 1 & tilskuere < 4000 ~ 1,
  nyt_hold == 1 & tilskuere > 4000 ~ 2))

# NA'erne i sidste_møde_tilskuere tildeles en ny værdi på baggrund af gennemsnitlige tilskuere for de t
```

```r
vff_all <- vff_all |>
  group_by(lille_stor_ny) |>
  mutate(avg_tilskuere = mean(tilskuere)) |>
  ungroup() |>
  mutate(
    sidste_møde_tilskuere = if_else(is.na(sidste_møde_tilskuere) & lille_stor_ny == 1, avg_tilskuere, s
    sidste_møde_tilskuere = if_else(is.na(sidste_møde_tilskuere) & lille_stor_ny == 2, avg_tilskuere, s
  )

# de irrelevante variabler der blev brugt til at udfylde NA'er slettes igen

vff_all <- vff_all |>
  select(-(nyt_hold:avg_tilskuere))

# for at fikse sidste_møde_resultat ændres NA observationerne til "første møde"

vff_all <- vff_all |>
  mutate(sidste_møde_resultat = if_else(is.na(sidste_møde_resultat), "første møde", sidste_møde_resulta

# NA'er i helligdage ændres til at hedde ingen, og observationer med manglende vejr data slettes

vff_all <- vff_all |>
  mutate(
    helligdag = if_else(is.na(helligdag), "ingen", helligdag)
  ) |>
  filter(!is.na(precip_past1h))

# ny variabel for om det er sommerferie eller ej og efterårsferie eller ej

vff_all <- vff_all |>
  mutate(
  sommerferie = as.factor(if_else(month(datotid) == 07 | (day(datotid) %in% 26:30 & month(datotid) == 0
)

# sæson_år variablen står stadig som en karakter, og for at prøve at fange sæson specifikke
# effekte, vil denne gerne beholdes som en faktor
```

```r
vff_all <- vff_all |>
  mutate(sæson_år = as.factor(sæson_år))

# for ikke at opfange for meget overlap mellem datotids variablen og sæsonåret,
# trækkes relevante variabler ud fra variablen, som nye variabler i stedet for og datotid fjernes

vff_all <- vff_all |>
  mutate(
    måned = month(datotid),
    uge_nr = week(datotid)
  ) |>
  select(-datotid)

# til sidst laves alle de kategoriske variabler til faktorer, og datasættet gemmes derefter
# som en rds fil

vff_all <- vff_all |>
  mutate(across(where(is.character), as.factor))

write_rds(vff_all, "data/vff_all.rds")
```

F Bilag 6: Modellering

F.1 Opsætning og indlæsning

Nødvendige pakker til modellering indlæses, og det processerede datasæt hentes fra RDS-filen.

```
# prerequisites ------------------------------------------------------------
# load packages

pacman::p_load(tidyverse, rvest, janitor, RSQLite, slider, leaps, glmnet)


# indlæser data fra rds fil

vff_all <- read_rds("data/vff_all.rds")


# for at sørge for der ikke opstår problemer med factor levels, som findes i testsættet
# men ikke i træningssættet, laves helligdagsvariablen om til om der blot var en hvilken som helst
# helligdag på dagen, eller ej. derefter grupperes de levels i modstander variablen, der kun optræder
# en enkelt gang, sådan at de får et fælles niveau

vff_all <- vff_all |>
  mutate(
    er_helligdag = as.factor(if_else(helligdag == "ingen", 0, 1)),
    modstander = fct_lump_min(modstander, min = 2, other_level = "andre_hold")
  ) |>
  select(-helligdag)
```

F.2 Opdeling i trænings- og testsæt

Datasættet opdeles i trænings- og testsæt (70/30) og derefter i fire forskellige versioner baseret på tidspunkt før kampen.

```
# sætter et seed

set.seed(2)

# opdeler dataen i træningssæt og testsæt

train_size <- floor(0.7 * nrow(vff_all))

train_data <- sample(nrow(vff_all), size = train_size)

test_data <- -train_data

vff_train <- vff_all[train_data, ]

vff_test <- vff_all[test_data, ]

# opdeler datasættene, sådan at vi får de datasæt vi skal bruge
# for at forudsige 1 måned før, 10 dage før, 7 dage før og 3 dage før

vff_train_1m <- vff_train |>
  select(-d10_tilskuere, -d7_tilskuere, -d3_tilskuere)

vff_train_d10 <- vff_train |>
  select(-d7_tilskuere, -d3_tilskuere)

vff_train_d7 <- vff_train |>
  select(-d10_tilskuere, -d3_tilskuere)

vff_train_d3 <- vff_train |>
  select(-d10_tilskuere, -d7_tilskuere)

vff_test_1m <- vff_test |>
  select(-d10_tilskuere, -d7_tilskuere, -d3_tilskuere)
```

```r
vff_test_d10 <- vff_test |>
  select(-d7_tilskuere, -d3_tilskuere)

vff_test_d7 <- vff_test |>
  select(-d10_tilskuere, -d3_tilskuere)

vff_test_d3 <- vff_test |>
  select(-d10_tilskuere, -d7_tilskuere)

# opretter en funktion til at lave de predictede værdier

predict.regsubsets <- function(object, newdata, id, ...) {
  form <- as.formula(object$call[[2]])
  mat <- model.matrix(form, newdata)
  coefi <- coef(object, id = id)
  xvars <- names(coefi)
  mat[, xvars] %*% coefi
}
```

F.3 Fuld lineær model

En fuld lineær regressionsmodel trænes for hvert af de fire tidspunkter (1 måned, 10 dage, 7 dage og 3 dage før kampen).

F.3.1 1 måned før kampstart

```r
# fuld linear model ----------------------------------------------------

lm_full_1m <- lm(tilskuere ~ ., vff_train_1m)

# finder predictede værdier

pred_full_1m <- predict(lm_full_1m,
                        vff_test_1m)

# test MSE
```

```
testmse_full_1m <- mean((vff_test_1m$tilskuere - pred_full_1m)^2)
```

F.3.2 10 dage før kampstart

```
## 10 dage inden kampstart ------------------------------------------------

lm_full_d10 <- lm(tilskuere ~ ., vff_train_d10)

# finder de predictede værdier

pred_full_d10 <- predict(lm_full_d10,
                         vff_test_d10)

# test MSE

testmse_full_d10 <- mean((vff_test_d10$tilskuere - pred_full_d10)^2)
```

F.3.3 7 dage før kampstart

```
## 7 dage inden kampstart -------------------------------------------------

lm_full_d7 <- lm(tilskuere ~ ., vff_train_d7)

# finder de predictede værdier

pred_full_d7 <- predict(lm_full_d7,
                        vff_test_d7)

# test MSE

testmse_full_d7 <- mean((vff_test_d7$tilskuere - pred_full_d7)^2)
```

F.3.4 3 dage før kampstart

```
## 3 dage inden kampstart -------------------------------------------------

lm_full_d3 <- lm(tilskuere ~ ., vff_train_d3)
```

```r
# finder de predictede værdier

predict_full_d3 <- predict(lm_full_d3,
                           vff_test_d3)

# test MSE

testmse_full_d3 <- mean((vff_test_d3$tilskuere - predict_full_d3)^2)
```

F.4 Forward selection

Forward stepwise selection bruges til at finde den optimale kombination af prædiktorer ved hjælp af 10-fold cross-validation.

```r
# forward selection ------------------------------------------------------

k <- 10 # Vi danner 10 folds
n <- nrow(vff_train) # registrerer hvor mange observationer, vi har.

folds <- sample(rep(1:k, length = n)) #Vi tildeler en værdi mellem 1 og
```

F.4.1 1 måned før kampstart

```r
## model for 1 måned før kampstart ---------------------------------------------

# laver en model matrix, for at finde nvmax, altså den maksimale variabelstørrele på kandidatmodellerne

model_matrix <- model.matrix(tilskuere ~ ., data = vff_train_1m)[, -1]

nvmax <- ncol(model_matrix)

cv.errors <- matrix(NA, k, nvmax,
                    dimnames = list(NULL, paste(1:nvmax)))

for (j in 1:k) {
  best.fit <- regsubsets(tilskuere ~ .,
                         data = vff_train_1m[folds != j, ],
                         nvmax = nvmax,
```

```
                          method = "forward")

    n_models <- length(summary(best.fit)$which[,1])
# pga. multikollinnearitet laves der ikke det samme
# antal modeller i hver fold, så vi finder det reelle antal der laves i hver fold og bruger det i næste
# af loopet

  for (i in 1:n_models) {
    pred <- predict(best.fit, vff_train_1m[folds == j, ], id = i)
    cv.errors[j, i] <- mean((vff_train$tilskuere[folds == j] - pred)^2)
  }
}

mean.cv.errors <- apply(cv.errors, 2, mean)

mean.cv.errors # vi får færre end 51 kandidatmodeller pga. kollinearitet

best_nvars_1m_fwd <- which.min(mean.cv.errors)

best_fwd_1m <- regsubsets(tilskuere ~ .,
                          data = vff_train_1m,
                          nvmax = nvmax,
                          method = "forward")

pred_fwd_1m <- predict(best_fwd_1m, vff_test_1m, id = best_nvars_1m_fwd)

testmse_fwd_1m <- mean((vff_test_1m$tilskuere - pred_fwd_1m)^2)
```

F.4.2 10 dage før kampstart

```
## model for 10 dage før kampstart ------------------------------------------------

# laver en model matrix, for at finde nvmax, altså den maksimale variabelstørrele på kandidatmodellerne

model_matrix <- model.matrix(tilskuere ~ ., data = vff_train_d10)[ , -1]

nvmax <- ncol(model_matrix)
```

```r
cv.errors <- matrix(NA, k, nvmax,
                    dimnames = list(NULL, paste(1:nvmax)))

for (j in 1:k) {
  best.fit <- regsubsets(tilskuere ~ .,
                         data = vff_train_d10[folds != j, ],
                         nvmax = nvmax,
                         method = "forward")

  n_models <- length(summary(best.fit)$which[,1])

  for (i in 1:n_models) {
    pred <- predict(best.fit, vff_train_d10[folds == j, ], id = i)
    cv.errors[j, i] <- mean((vff_train$tilskuere[folds == j] - pred)^2)
  }
}

mean.cv.errors <- apply(cv.errors, 2, mean)

mean.cv.errors

best_nvars_d10_fwd <- which.min(mean.cv.errors)

best_fwd_d10 <- regsubsets(tilskuere ~ .,
                         data = vff_train_d10,
                         nvmax = nvmax,
                         method = "forward")

pred_fwd_d10 <- predict(best_fwd_d10, vff_test_d10, id = best_nvars_d10_fwd)

testmse_fwd_d10 <- mean((vff_test_d10$tilskuere - pred_fwd_d10)^2)
```

F.4.3 7 dage før kampstart

```r
## model for 7 dage før kampstart ------------------------------------------------

# laver en model matrix, for at finde nvmax, altså den maksimale variabelstørrele på kandidatmodellerne
```

```r
model_matrix <- model.matrix(tilskuere ~ ., data = vff_train_d7)[ , -1]

nvmax <- ncol(model_matrix)

cv.errors <- matrix(NA, k, nvmax,
                    dimnames = list(NULL, paste(1:nvmax)))

for (j in 1:k) {
  best.fit <- regsubsets(tilskuere ~ .,
                         data = vff_train_d7[folds != j, ],
                         nvmax = nvmax,
                         method = "forward")

  n_models <- length(summary(best.fit)$which[,1])

  for (i in 1:n_models) {
    pred <- predict(best.fit, vff_train_d7[folds == j, ], id = i)
    cv.errors[j, i] <- mean((vff_train$tilskuere[folds == j] - pred)^2)
  }
}

mean.cv.errors <- apply(cv.errors, 2, mean)

mean.cv.errors

best_nvars_d7_fwd <- which.min(mean.cv.errors)

best_fwd_d7 <- regsubsets(tilskuere ~ .,
                         data = vff_train_d7,
                         nvmax = nvmax,
                         method = "forward")

pred_fwd_d7 <- predict(best_fwd_d7, vff_test_d7, id = best_nvars_d7_fwd)

testmse_fwd_d7 <- mean((vff_test_d7$tilskuere - pred_fwd_d7)^2)
```

F.4.4 3 dage før kampstart

```
## model for 3 dage før kampstart ------------------------------------------------

# laver en model matrix, for at finde nvmax, altså den maksimale variabelstørrele på kandidatmodellerne

model_matrix <- model.matrix(tilskuere ~ ., data = vff_train_d3)[ , -1]

nvmax <- ncol(model_matrix)

cv.errors <- matrix(NA, k, nvmax,
                    dimnames = list(NULL, paste(1:nvmax)))

for (j in 1:k) {
  best.fit <- regsubsets(tilskuere ~ .,
                         data = vff_train_d3[folds != j, ],
                         nvmax = nvmax,
                         method = "forward")

  n_models <- length(summary(best.fit)$which[,1])

  for (i in 1:n_models) {
    pred <- predict(best.fit, vff_train_d3[folds == j, ], id = i)
    cv.errors[j, i] <- mean((vff_train$tilskuere[folds == j] - pred)^2)
  }
}

mean.cv.errors <- apply(cv.errors, 2, mean)

mean.cv.errors

best_nvars_d3_fwd <- which.min(mean.cv.errors)

best_fwd_d3 <- regsubsets(tilskuere ~ .,
                         data = vff_train_d3,
                         nvmax = nvmax,
                         method = "forward")
```

```r
pred_fwd_d3 <- predict(best_fwd_d3, vff_test_d3, id = best_nvars_d3_fwd)

testmse_fwd_d3 <- mean((vff_test_d3$tilskuere - pred_fwd_d3)^2)
```

F.5 Backward selection

Backward stepwise selection bruges som alternativ metode til at finde den optimale kombination af prædiktorer ved hjælp af 10-fold cross-validation.

```r
# backward selection ------------------------------------------------------

k <- 10 # Vi danner 10 folds
n <- nrow(vff_train) # registrerer hvor mange observationer, vi har.

folds <- sample(rep(1:k, length = n)) #Vi tildeler en værdi mellem 1 og
```

F.5.1 1 måned før kampstart

```r
## model for 1 måned før kampstart -----------------------------------------------

# laver en model matrix, for at finde nvmax, altså den maksimale variabelstørrele på kandidatmodellerne

model_matrix <- model.matrix(tilskuere ~ ., data = vff_train_1m)[, -1]

nvmax <- ncol(model_matrix)

cv.errors <- matrix(NA, k, nvmax,
                    dimnames = list(NULL, paste(1:nvmax)))

for (j in 1:k) {
  best.fit <- regsubsets(tilskuere ~ .,
                         data = vff_train_1m[folds != j, ],
                         nvmax = nvmax,
                         method = "backward")

    n_models <- length(summary(best.fit)$which[,1])

  for (i in 1:n_models) {
```

```
    pred <- predict(best.fit, vff_train_1m[folds == j, ], id = i)
    cv.errors[j, i] <- mean((vff_train$tilskuere[folds == j] - pred)^2)
  }
}

mean.cv.errors <- apply(cv.errors, 2, mean)

mean.cv.errors

best_nvars_1m_bwd <- which.min(mean.cv.errors)

best_bwd_1m <- regsubsets(tilskuere ~ .,
                          data = vff_train_1m,
                          nvmax = nvmax,
                          method = "backward")

coef(best_bwd_1m, best_nvars_1m_bwd)

pred_bwd_1m <- predict(best_bwd_1m, vff_test_1m, id = best_nvars_1m_bwd)

testmse_bwd_1m <- mean((vff_test_1m$tilskuere - pred_bwd_1m)^2)
```

F.5.2 10 dage før kampstart

```
## model for 10 dage før kampstart ------------------------------------------------

# laver en model matrix, for at finde nvmax, altså den maksimale variabelstørrele på kandidatmodellerne

model_matrix <- model.matrix(tilskuere ~ ., data = vff_train_d10)[ , -1]

nvmax <- ncol(model_matrix)

cv.errors <- matrix(NA, k, nvmax,
                    dimnames = list(NULL, paste(1:nvmax)))

for (j in 1:k) {
  best.fit <- regsubsets(tilskuere ~ .,
```

```
                          data = vff_train_d10[folds != j, ],
                          nvmax = nvmax,
                          method = "backward")

  n_models <- length(summary(best.fit)$which[,1])

  for (i in 1:n_models) {
    pred <- predict(best.fit, vff_train_d10[folds == j, ], id = i)
    cv.errors[j, i] <- mean((vff_train$tilskuere[folds == j] - pred)^2)
  }
}

mean.cv.errors <- apply(cv.errors, 2, mean)

mean.cv.errors

best_nvars_d10_bwd <- which.min(mean.cv.errors)

best_bwd_d10 <- regsubsets(tilskuere ~ .,
                          data = vff_train_d10,
                          nvmax = nvmax,
                          method = "backward")

pred_bwd_d10 <- predict(best_bwd_d10, vff_test_d10, id = best_nvars_d10_bwd)

testmse_bwd_d10 <- mean((vff_test_d10$tilskuere - pred_bwd_d10)^2)
```

F.5.3 7 dage før kampstart

```
## model for 7 dage før kampstart ------------------------------------------------

# laver en model matrix, for at finde nvmax, altså den maksimale variabelstørrele på kandidatmodellerne

model_matrix <- model.matrix(tilskuere ~ ., data = vff_train_d7)[ , -1]

nvmax <- ncol(model_matrix)
```

```
cv.errors <- matrix(NA, k, nvmax,
                    dimnames = list(NULL, paste(1:nvmax)))

for (j in 1:k) {
  best.fit <- regsubsets(tilskuere ~ .,
                         data = vff_train_d7[folds != j, ],
                         nvmax = nvmax,
                         method = "backward")

  n_models <- length(summary(best.fit)$which[,1])

  for (i in 1:n_models) {
    pred <- predict(best.fit, vff_train_d7[folds == j, ], id = i)
    cv.errors[j, i] <- mean((vff_train$tilskuere[folds == j] - pred)^2)
  }
}

mean.cv.errors <- apply(cv.errors, 2, mean)

mean.cv.errors

best_nvars_d7_bwd <- which.min(mean.cv.errors)

best_bwd_d7 <- regsubsets(tilskuere ~ .,
                         data = vff_train_d7,
                         nvmax = nvmax,
                         method = "backward")

pred_bwd_d7 <- predict(best_bwd_d7, vff_test_d7, id = best_nvars_d7_bwd)

testmse_bwd_d7 <- mean((vff_test_d7$tilskuere - pred_bwd_d7)^2)
```

F.5.4 3 dage før kampstart

```
## model for 3 dage før kampstart ------------------------------------------------

# laver en model matrix, for at finde nvmax, altså den maksimale variabelstørrele på kandidatmodellerne
```

```r
model_matrix <- model.matrix(tilskuere ~ ., data = vff_train_d3)[ , -1]

nvmax <- ncol(model_matrix)

cv.errors <- matrix(NA, k, nvmax,
                    dimnames = list(NULL, paste(1:nvmax)))

for (j in 1:k) {
  best.fit <- regsubsets(tilskuere ~ .,
                         data = vff_train_d3[folds != j, ],
                         nvmax = nvmax,
                         method = "backward")

  n_models <- length(summary(best.fit)$which[,1])

  for (i in 1:n_models) {
    pred <- predict(best.fit, vff_train_d3[folds == j, ], id = i)
    cv.errors[j, i] <- mean((vff_train$tilskuere[folds == j] - pred)^2)
  }
}

mean.cv.errors <- apply(cv.errors, 2, mean)

mean.cv.errors

best_nvars_d3_bwd <- which.min(mean.cv.errors)

best_bwd_d3 <- regsubsets(tilskuere ~ .,
                         data = vff_train_d3,
                         nvmax = nvmax,
                         method = "backward")

pred_bwd_d3 <- predict(best_bwd_d3, vff_test_d3, id = best_nvars_d3_bwd)

testmse_bwd_d3 <- mean((vff_test_d3$tilskuere - pred_bwd_d3)^2)
```

F.6 Ridge & Lasso regression

Ridge regression anvendes til at håndtere multikollinearitet ved at tilføje en L2-penalty til modellen. Lambda-parameteren optimeres via 10-fold cross-validation, mens Lasso regression anvendes til både variable selection og regularisering ved at tilføje en L1-penalty. Lambda optimeres via 10-fold cross-validation.

F.6.1 1 måned før kampstart

```
# ridge og lasso regression ------------------------------------------------------------

## modeller for 1 måned inden kampen ----------------------------------------

# opretter objekter for x variablerne og y variablen. x variablerne sættes i en matrice,
# mens y blot er outputtet, her antal tilskuere

x_train_1m <- model.matrix(tilskuere ~ ., data = vff_train_1m)[, -1]

y_train_1m <- vff_train_1m$tilskuere

x_test_1m <- model.matrix(tilskuere ~ ., data = vff_test_1m)[, -1]
```

```
### ridge regression ------------------------------------------------------

# optimerer tuning parameteren med k-fold cross validation (10 folds), for at finde den bedste lambda

cv_ridge_1m <- cv.glmnet(x_train_1m, y_train_1m, alpha = 0, nfolds = 10)

# bedste lambda

bestlambda_ridge_1m <- cv_ridge_1m$lambda.min

# bedste lambda indenfor 1 standardafvigelse af den optimale lambda (simpleste model)

ridgelambda_1se_1m <- cv_ridge_1m$lambda.1se

# endelige model

final_ridge_1m <- glmnet(x_train_1m, y_train_1m, alpha = 0, lambda = bestlambda_ridge_1m)
```

```r
# modellens koefficienter

ridge_coefs_1m <- coef(final_ridge_1m)

# predictede værdier

ridge_pred_1m <- predict(final_ridge_1m, s = bestlambda_ridge_1m, newx = x_test_1m)

# test MSE

testmse_ridge_1m <- mean((vff_test_1m$tilskuere - ridge_pred_1m)^2)
```

```r
### lasso regression ----------------------------------------------------

# optimerer tuning parameteren med k-fold cross validation (10 folds), for at finde den bedste lambda

cv_lasso_1m <- cv.glmnet(x_train_1m, y_train_1m, alpha = 1, nfolds = 10)

# bedste lambda

bestlambda_lasso_1m <- cv_lasso_1m$lambda.min

# bedste lambda indenfor 1 standardafvigelse af den optimale lambda (simpleste model)

lassolambda_1se_1m <- cv_lasso_1m$lambda.1se

# endelige model

final_lasso_1m <- glmnet(x_train_1m, y_train_1m, alpha = 1, lambda = bestlambda_lasso_1m)

# modellens koefficienter

lasso_coefs_1m <- coef(final_lasso_1m)

# predictede værdier

lasso_pred_1m <- predict(final_lasso_1m, s = bestlambda_lasso_1m, newx = x_test_1m)
```

```r
# test MSE

testmse_lasso_1m <- mean((vff_test_1m$tilskuere - lasso_pred_1m)^2)
```

F.6.2 10 dage før kampstart

```r
## modeller for 10 dage inden kampen ----------------------------------------

# opretter objekter for x variablerne og y variablen. x variablerne sættes i en matrice,
# mens y blot er outputtet, her antal tilskuere

x_train_d10 <- model.matrix(tilskuere ~ ., data = vff_train_d10)[, -1]

y_train_d10 <- vff_train_d10$tilskuere

x_test_d10 <- model.matrix(tilskuere ~ ., data = vff_test_d10)[, -1]
```

```r
### ridge regression ------------------------------------------------------

# optimerer tuning parameteren med k-fold cross validation (10 folds), for at finde den bedste lambda

cv_ridge_d10 <- cv.glmnet(x_train_d10, y_train_d10, alpha = 0, nfolds = 10)

# bedste lambda

bestlambda_ridge_d10 <- cv_ridge_d10$lambda.min

# bedste lambda indenfor 1 standardafvigelse af den optimale lambda (simpleste model)

ridgelambda_1se_d10 <- cv_ridge_d10$lambda.1se

# endelige model

final_ridge_d10 <- glmnet(x_train_d10, y_train_d10, alpha = 0, lambda = bestlambda_ridge_d10)

# modellens koefficienter
```

```r
ridge_coefs_d10 <- coef(final_ridge_d10)

# predictede værdier

ridge_pred_d10 <- predict(final_ridge_d10, s = bestlambda_ridge_d10, newx = x_test_d10)

# test MSE

testmse_ridge_d10 <- mean((vff_test_d10$tilskuere - ridge_pred_d10)^2)
```

```r
### lasso regression ------------------------------------------------------

# optimerer tuning parameteren med k-fold cross validation (10 folds), for at finde den bedste lambda

cv_lasso_d10 <- cv.glmnet(x_train_d10, y_train_d10, alpha = 1, nfolds = 10)

# bedste lambda

bestlambda_lasso_d10 <- cv_lasso_d10$lambda.min

# bedste lambda indenfor 1 standardafvigelse af den optimale lambda (simpleste model)

lassolambda_1se_d10 <- cv_lasso_d10$lambda.1se

# endelige model

final_lasso_d10 <- glmnet(x_train_d10, y_train_d10, alpha = 1, lambda = bestlambda_lasso_d10)

# modellens koefficienter

lasso_coefs_d10 <- coef(final_lasso_d10)

# predictede værdier

lasso_pred_d10 <- predict(final_lasso_d10, s = bestlambda_lasso_d10, newx = x_test_d10)

# test MSE
```

```r
testmse_lasso_d10 <- mean((vff_test_d10$tilskuere - lasso_pred_d10)^2)
```

F.6.3 7 dage før kampstart

```r
## modeller for 7 dage inden kampen ----------------------------------------

# opretter objekter for x variablerne og y variablen. x variablerne sættes i en matrice,
# mens y blot er outputtet, her antal tilskuere

x_train_d7 <- model.matrix(tilskuere ~ ., data = vff_train_d7)[, -1]

y_train_d7 <- vff_train_d7$tilskuere

x_test_d7 <- model.matrix(tilskuere ~ ., data = vff_test_d7)[, -1]
```

```r
### ridge regression ------------------------------------------------------

# optimerer tuning parameteren med k-fold cross validation (10 folds), for at finde den bedste lambda

cv_ridge_d7 <- cv.glmnet(x_train_d7, y_train_d7, alpha = 0, nfolds = 10)

# bedste lambda

bestlambda_ridge_d7 <- cv_ridge_d7$lambda.min

# bedste lambda indenfor 1 standardafvigelse af den optimale lambda (simpleste model)

ridgelambda_1se_d7 <- cv_ridge_d7$lambda.1se

# endelige model

final_ridge_d7 <- glmnet(x_train_d7, y_train_d7, alpha = 0, lambda = bestlambda_ridge_d7)

# modellens koefficienter

ridge_coefs_d7 <- coef(final_ridge_d7)
```

```r
# predictede værdier

ridge_pred_d7 <- predict(final_ridge_d7, s = bestlambda_ridge_d7, newx = x_test_d7)

# test MSE

testmse_ridge_d7 <- mean((vff_test_d7$tilskuere - ridge_pred_d7)^2)
```

```r
### lasso regression ------------------------------------------------------

# optimerer tuning parameteren med k-fold cross validation (10 folds), for at finde den bedste lambda

cv_lasso_d7 <- cv.glmnet(x_train_d7, y_train_d7, alpha = 1, nfolds = 10)

# bedste lambda

bestlambda_lasso_d7 <- cv_lasso_d7$lambda.min

# bedste lambda indenfor 1 standardafvigelse af den optimale lambda (simpleste model)

lassolambda_1se_d7 <- cv_lasso_d7$lambda.1se

# endelige model

final_lasso_d7 <- glmnet(x_train_d7, y_train_d7, alpha = 1, lambda = bestlambda_lasso_d7)

# modellens koefficienter

lasso_coefs_d7 <- coef(final_lasso_d7)

# predictede værdier

lasso_pred_d7 <- predict(final_lasso_d7, s = bestlambda_lasso_d7, newx = x_test_d7)

# test MSE

testmse_lasso_d7 <- mean((vff_test_d7$tilskuere - lasso_pred_d7)^2)
```

F.6.4 3 dage før kampstart

```
## modeller for 3 dage inden kampen ---------------------------------------

# opretter objekter for x variablerne og y variablen. x variablerne sættes i en matrice,
# mens y blot er outputtet, her antal tilskuere

x_train_d3 <- model.matrix(tilskuere ~ ., data = vff_train_d3)[, -1]

y_train_d3 <- vff_train_d3$tilskuere

x_test_d3 <- model.matrix(tilskuere ~ ., data = vff_test_d3)[, -1]

### ridge regression ------------------------------------------------------

# optimerer tuning parameteren med k-fold cross validation (10 folds), for at finde den bedste lambda

cv_ridge_d3 <- cv.glmnet(x_train_d3, y_train_d3, alpha = 0, nfolds = 10)

# bedste lambda

bestlambda_ridge_d3 <- cv_ridge_d3$lambda.min

# bedste lambda indenfor 1 standardafvigelse af den optimale lambda (simpleste model)

ridgelambda_1se_d3 <- cv_ridge_d3$lambda.1se

# endelige model

final_ridge_d3 <- glmnet(x_train_d3, y_train_d3, alpha = 0, lambda = bestlambda_ridge_d3)

# modellens koefficienter

ridge_coefs_d3 <- coef(final_ridge_d3)

# predictede værdier
```

```r
ridge_pred_d3 <- predict(final_ridge_d3, s = bestlambda_ridge_d3, newx = x_test_d3)

# test MSE

testmse_ridge_d3 <- mean((vff_test_d3$tilskuere - ridge_pred_d3)^2)
```

```r
### lasso regression ----------------------------------------------------

# optimerer tuning parameteren med k-fold cross validation (10 folds), for at finde den bedste lambda

cv_lasso_d3 <- cv.glmnet(x_train_d3, y_train_d3, alpha = 1, nfolds = 10)

# bedste lambda

bestlambda_lasso_d3 <- cv_lasso_d3$lambda.min

# bedste lambda indenfor 1 standardafvigelse af den optimale lambda (simpleste model)

lassolambda_1se_d3 <- cv_lasso_d3$lambda.1se

# endelige model

final_lasso_d3 <- glmnet(x_train_d3, y_train_d3, alpha = 1, lambda = bestlambda_lasso_d3)

# modellens koefficienter

lasso_coefs_d3 <- coef(final_lasso_d3)

# predictede værdier

lasso_pred_d3 <- predict(final_lasso_d3, s = bestlambda_lasso_d3, newx = x_test_d3)

# test MSE

testmse_lasso_d3 <- mean((vff_test_d3$tilskuere - lasso_pred_d3)^2)
```

F.7 Visualisering af MSE & RMSE

MSE og RMSE visualiseres for alle modeller på tværs af de fire prædiktionstidspunkter, for at identificere den bedst performende model.

```
# visualiseringer af MSE og RMSE ----------------------------------------

# laver dataframes med MSE og RMSE for alle modellerne
test_mse_df <- data.frame(
  måned1 = c(testmse_full_1m, testmse_fwd_1m, testmse_bwd_1m, testmse_ridge_1m, testmse_lasso_1m),
  dag10 = c(testmse_full_d10, testmse_fwd_d10, testmse_bwd_d10, testmse_ridge_d10, testmse_lasso_d10),
  dag7 = c(testmse_full_d7, testmse_fwd_d7, testmse_bwd_d7, testmse_ridge_d7, testmse_lasso_d7),
  dag3 = c(testmse_full_d3, testmse_fwd_d3, testmse_bwd_d3, testmse_ridge_d3, testmse_lasso_d3),
  row.names = c("Full Linear", "Forward Selection", "Backward Selection",
                "Ridge", "Lasso")
)

test_rmse_df <- data.frame(
  måned1 = c(sqrt(testmse_full_1m), sqrt(testmse_fwd_1m), sqrt(testmse_bwd_1m), sqrt(testmse_ridge_1m),
  dag10 = c(sqrt(testmse_full_d10), sqrt(testmse_fwd_d10), sqrt(testmse_bwd_d10), sqrt(testmse_ridge_d1
  dag7 = c(sqrt(testmse_full_d7), sqrt(testmse_fwd_d7), sqrt(testmse_bwd_d7), sqrt(testmse_ridge_d7), s
  dag3 = c(sqrt(testmse_full_d3), sqrt(testmse_fwd_d3), sqrt(testmse_bwd_d3), sqrt(testmse_ridge_d3), s
  row.names = c("Full Linear", "Forward Selection", "Backward Selection",
                "Ridge", "Lasso")
)

best_models <- data.frame(
  best_model = apply(test_mse_df, 2, function(x) rownames(test_mse_df)[which.min(x)]),
  MSE = apply(test_mse_df, 2, min)
) |>
  mutate(RMSE = sqrt(MSE))

best_models

# laver plots over sammenligninger på modellerne

test_rmse_long <- test_rmse_df |>
  rownames_to_column("Model") |>
```

```r
  pivot_longer(cols = -Model, names_to = "Timeframe", values_to = "RMSE")

test_rmse_long <- test_rmse_long |>
  mutate(
    Timeframe = factor(
      Timeframe,
      levels = c("måned1", "dag10", "dag7", "dag3")
    ))

ggplot(test_rmse_long, aes(x = RMSE, y = reorder(Model, -RMSE), fill = Model)) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = round(RMSE, 0)), hjust = -0.2, size = 3.5) +
  facet_wrap(~Timeframe, scales = "free_x", ncol = 2,
             labeller = labeller(Timeframe = c("dag3" = "3 Dage", "dag7" = "7 Dage",
                                               "dag10" = "10 Dage", "måned1" = "1 Måned"))) +
  scale_fill_brewer(palette = "Set2") +
  labs(title = "Model Sammenligning Over Forskellige Prediction Tidspunkter",
       x = "Root Mean Squared Error (RMSE)",
       y = NULL) +
  theme_minimal(base_size = 12) +
  theme(plot.title = element_text(face = "bold"),
        strip.text = element_text(face = "bold", size = 12))

test_rmse_long_best <- test_rmse_long |>
  filter((Model == "Ridge" & Timeframe == "måned1") | (Model == "Lasso" & Timeframe == "dag10") |
         (Model == "Lasso" & Timeframe == "dag7") | (Model == "Lasso" & Timeframe == "dag3"))

test_rmse_long_best <- test_rmse_long_best |>
    mutate(
    Timeframe = factor(Timeframe,
                       levels = c("måned1", "dag10", "dag7", "dag3"))
  )

test_rmse_long_best <- test_rmse_long_best |>
  mutate(
    Timeframe = recode(
      Timeframe,
```

```
      "dag3"    = "3 Dage",
      "dag7"    = "7 Dage",
      "dag10"   = "10 Dage",
      "måned1"  = "1 Måned"
    ),
    Timeframe = factor(
      Timeframe,
      levels = c("1 Måned", "10 Dage", "7 Dage", "3 Dage")
    )
  )

ggplot(test_rmse_long_best,
       aes(x = Timeframe, y = RMSE, fill = Model)) +
  geom_col(width = 0.6) +
  coord_flip() +
  geom_text(aes(label = round(RMSE, 0)), hjust = -0.1, size = 5) +
  labs(
    title = "Bedst performende ML-modeller",
    subtitle = "Sammenligning på tværs af tid før kampstart",
    x = "Tid før kampstart",
    y = "RMSE",
    fill = "Model"
  ) +
  theme_minimal(base_size = 12) +
  theme(plot.title = element_text(face = "bold")) +
  scale_fill_brewer(palette = "Set2")
```

F.8 Nye prædiktioner

Data for den kommende kamp klargøres med alle nødvendige prædiktorer i det korrekte format. De bedste modeller anvendes til at forudsige tilskuertal for den kommende kamp på tværs af alle fire prædiktionstidspunkter.

```
# nye predictions ------------------------------------------------------

# nyt datasæt for den næste hjemmekampe i sæsonen mod BIF, som endnu ikke er spillet
# der er to observationer af den samme kamp, hvoraf den ene er et best case scenario
# og den anden er et worst case scenario

predictions_nye <- data.frame(
  ugedag = c("Søn", "Søn"),
  vff_resultat_lagged = c("uafgjort", "uafgjort"),
  mål_sidste_kamp = c(1, 1),
  sæson_år = c("2026", "2026"),
  runde_nr = c(20, 20),
  mål_sidste3_lagged = c(8, 3),
  point_sidste3_lagged = c(5, 2),
  point_sæson_lagged = c(27, 24),
  tidspunkt = c("aften", "aften"),
  sidste_møde_tilskuere = c(7658, 7658),
  modstander = c("BIF", "BIF"),
  sidste_møde_resultat = c("vundet", "vundet"),
  temp_dry = c(8, -2),
  wind_speed = c(0.1, 7.4),
  precip_past1h = c(0, 1.5),
  d10_tilskuere = c(3987, 2264),
  d7_tilskuere = c(5021, 3191),
  d3_tilskuere = c(6183, 3972),
  sommerferie = c(0, 0),
  måned = c(2, 2),
  uge_nr = c(7, 7),
  er_helligdag = c(0, 0)
)

# sørger for at alt er faktorer, og at faktorerne har samme levels som det originale datasæt
```

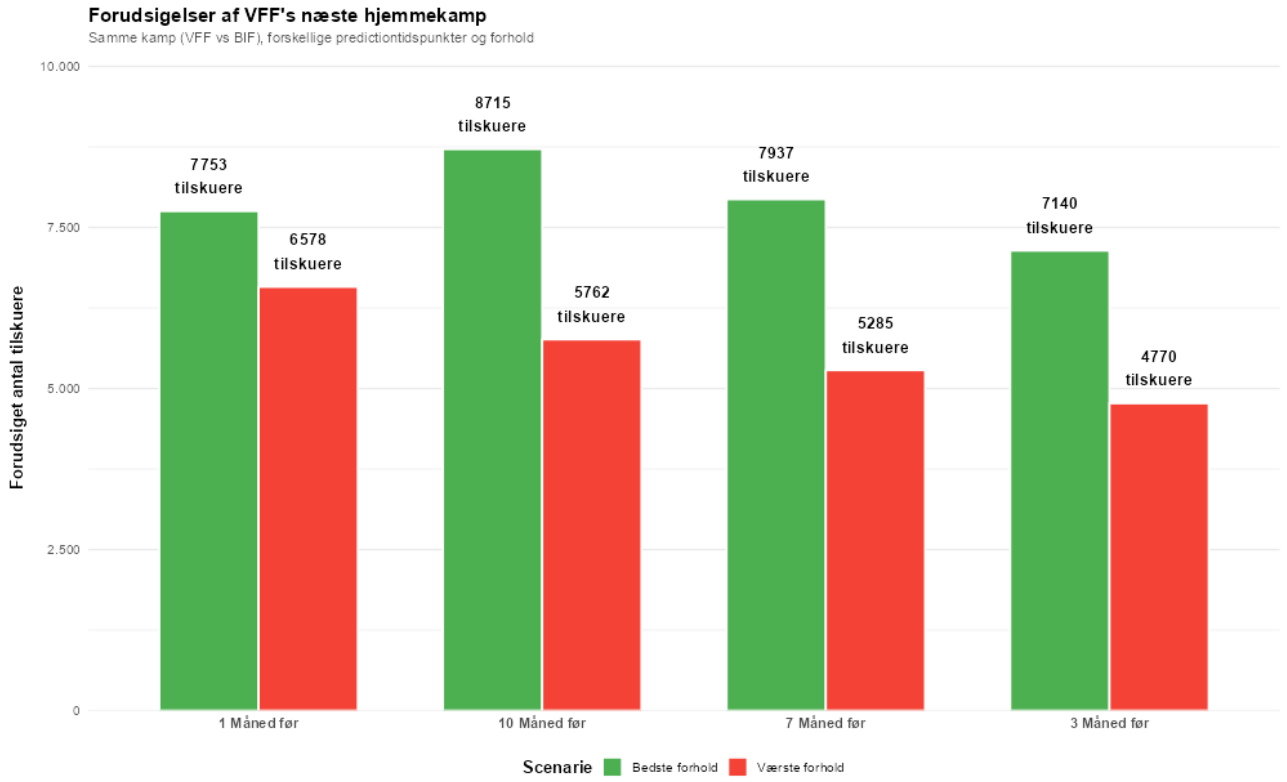
```r
predictions_nye <- predictions_nye |>
  mutate(
    ugedag = factor(ugedag, levels = levels(vff_all$ugedag)),
    vff_resultat_lagged = factor(vff_resultat_lagged, levels = levels(vff_all$vff_resultat_lagged)),
    sæson_år = factor(sæson_år, levels = levels(vff_all$sæson_år)),
    tidspunkt = factor(tidspunkt, levels = levels(vff_all$tidspunkt)),
    modstander = factor(modstander, levels = levels(vff_all$modstander)),
    sidste_møde_resultat = factor(sidste_møde_resultat, levels = levels(vff_all$sidste_møde_resultat)),
    sommerferie = factor(sommerferie, levels = levels(vff_all$sommerferie)),
    er_helligdag = factor(er_helligdag, levels = levels(vff_all$er_helligdag))
  )

# datasæt med de rigtige variabler for hvert tidspunkt før kampstart

nye_1m <- predictions_nye |>
  select(-d10_tilskuere, -d7_tilskuere, -d3_tilskuere)

nye_d10 <- predictions_nye |>
  select(-d7_tilskuere, -d3_tilskuere)

nye_d7 <- predictions_nye |>
  select(-d10_tilskuere, -d3_tilskuere)

nye_d3 <- predictions_nye |>
  select(-d10_tilskuere, -d7_tilskuere)

# predictions for den nye kamp, lavet på de modeller der performede bedst på hvert tidspunkt

x_nye_1m <- model.matrix(~ ., data = nye_1m)[, -1]
nye_pred_1m <- predict(final_ridge_1m, newx = x_nye_1m)

x_nye_d10 <- model.matrix(~ ., data = nye_d10)[, -1]
nye_pred_d10 <- predict(final_lasso_d10, newx = x_nye_d10)

x_nye_d7 <- model.matrix(~ ., data = nye_d7)[, -1]
nye_pred_d7 <- predict(final_lasso_d7, newx = x_nye_d7)
```

```r
x_nye_d3 <- model.matrix(~ ., data = nye_d3)[, -1]
nye_pred_d3 <- predict(final_lasso_d3, s = bestlambda_lasso_d3, newx = x_nye_d3)

# visualisering af de nye predictions
# laver en dataframe med resultaterne fra predictions

results <- data.frame(
  Scenario = c("Best Case", "Worst Case"),
  `1 Måned før` = c(nye_pred_1m[1], nye_pred_1m[2]),
  `10 Måned før` = c(nye_pred_d10[1], nye_pred_d10[2]),
  `7 Måned før` = c(nye_pred_d7[1], nye_pred_d7[2]),
  `3 Måned før` = c(nye_pred_d3[1], nye_pred_d3[2]),
  check.names = FALSE
)

results_long <- results |>
  pivot_longer(cols = -Scenario, names_to = "Timeframe", values_to = "Predicted_Attendance") |>
  mutate(Timeframe = factor(Timeframe, levels = c("1 Måned før", "10 Måned før",
                                                  "7 Måned før", "3 Måned før")))

results_long <- results_long |>
  mutate(
    days_before = case_when(
      Timeframe == "1 Måned før" ~ 30,
      Timeframe == "10 Måned før" ~ 10,
      Timeframe == "7 Måned før" ~ 7,
      Timeframe == "3 Måned før" ~ 3
    ),
    label_text = paste0(round(Predicted_Attendance, 0), " tilskuere\n(",
                        days_before, " dage før)")
  )

# plot af resultater

ggplot(results_long, aes(x = Timeframe, y = Predicted_Attendance, fill = Scenario)) +
  geom_col(position = "dodge", width = 0.7, color = "white", linewidth = 0.8) +
  geom_text(aes(label = paste0(round(Predicted_Attendance, 0), "\ntilskuere")),
```

```
          position = position_dodge(width = 0.7),
          vjust = -0.5, size = 4.5, fontface = "bold") +
scale_fill_manual(
  values = c("Best Case" = "#4CAF50", "Worst Case" = "#F44336"),
  labels = c("Bedste forhold", "Værste forhold")
) +
scale_y_continuous(expand = expansion(mult = c(0, 0.15)),
                   labels = function(x) format(x, big.mark = ".", decimal.mark = ",")) +
labs(title = "Forudsigelser af VFF's næste hjemmekamp",
     subtitle = "Samme kamp (VFF vs BIF), forskellige predictiontidspunkter og forhold",
     x = NULL,
     y = "Forudsiget antal tilskuere",
     fill = "Scenarie") +
theme_minimal(base_size = 13) +
theme(
  legend.position = "bottom",
  plot.title = element_text(face = "bold", size = 15, margin = margin(b = 5)),
  plot.subtitle = element_text(size = 11, color = "gray30", margin = margin(b = 15)),
  plot.caption = element_text(hjust = 0, face = "italic", color = "gray50"),
  axis.text.x = element_text(face = "bold", size = 11),
  axis.title.y = element_text(face = "bold", margin = margin(r = 10)),
  panel.grid.major.x = element_blank(),
  legend.title = element_text(face = "bold")
)
```

F.8.1 Resultater af prædiktioner

For at illustrere hvordan nye prædiktioner kan ændre sig med forskellige observationer og på forskellige prædiktionstidspunkter, har vi opstillet to scenarier for VFF's næste hjemmekamp mod BIF. Der er opstillet et "Best case scenario", hvor forholdene er gode, hvilket vil sige at der er godt vejr (8°C, ingen regn, lav vind), god sportslig form i de seneste kampe (8 mål og 5 point i de sidste tre kampe) og højt billetsalg i dagene op til kampen (6,183 billeter). Derudover er der også opstillet et "Worst case scenario", hvor forholdene er betydeligt dårligere, og der altså er dårligt vejr (-2°C, regn og høj vindstyrke), dårligere sportslig form i de seneste kampe (3 mål og 2 point i de sidste tre kampe) og lavt billetsalg i dagene op til kampen (3,972 billeter).



Figur 9: Sammenligning af nye prædiktioner

Plottet viser, at at prædiktionerne samlet set falder når man bevæger sig fra forudsigelser 1 måned inden kampstart til 3 dage før kampstart. Dette skal dog ikke tolkes som et fald i modellernes performance, men snarere som et udtryk for at prædiktionerne blivere mere præcise når mere informativ data inkluderes. Det er især inkluderingen af de faktiske billetsalgsdata tættere på kampdagene giver modellerne et stærkere grundlag for at forudsige tilskuertallene. Samtidig ses en voksende forskel i prædiktionerne mellem best case og worst case scenarierne, jo tættere man kommer på kampstart, hvilket igen kan forklares af inddragelsen af billetsalgsdataen. Ved forudsigelser måned før kampstart, indgår billetsalget endnu ikke i modellen, hvorfor prædiktionsforskel mellem de to scenarier kan forklares af variationer mellem vejrforholdene og den sportslige form. Når billetsalgsvariablerne derimod inkluderes, vokser prædiktionsforskellen mellem de to scenarier markant, hvilket indikerer at disse variabler har stor betydning for modellernes nøjagtighed. Når vi har et worst case scenarie hvor både billetsalget i

dagene op til kampen er dårligt og de øvrige forhold også er, vil der altså forudsiges et endnu dårligere tilskuertal fordi variablen er så værdifuld for modellen, og vice versa. Ud fra dette kan det konkluderes, at billetsalgsdata er en afgørende variabel for prædiktionernes nøjagtighed og de forskelle der opstår mellem scenarier tæt på kampstart. I en forretningsmæssig kontekst betyder dette, at modellen i højere grad kan anvendes som et operationelt beslutningsværktøj i dagene op til kamp, hvor den kan give mere realistiske skøn for de forventede tilskuertal. Denne viden kan anvendes i planlægningen af bemanding, indkøb, og markedsføringsindsatser. Samtidig kan en scenarietænkning som denne, give den kommercielle afdeling mulighed for at arbejde proaktivt med forskellige udfald og dermed være bedre forberedt i dagene op til hjemmekampene.