

Optimeret Bygningsanvendelse

Signaturprojekt – AI

William Klakk Jeppesen, Emil Bekker og Nicolai Thomsen
NTT DATA Business Solutions A/S

Table of contents

1.	Introduktion til projektet	6
2.	Infrastruktur	6
2.1	Titan	7
2.1.1	Azure Subscription / Tenant	7
2.1.2	API	8
2.1.3	Database	8
2.1.4	Frontend / Portal	8
2.2	Udviklet i projektet	9
2.2.1	Dataindsamling / Landing zone	9
2.2.1.1	Data Pipelines	9
2.2.2	Data storage	10
2.2.2.1	Storage Account	10
2.2.3	Frontend – Power Bi	10
2.2.4	AI-modellerne	10
3.	Transformationer	11
3.1	Fremgangsmåde	11
4.	Business Intelligence/Frontend	12
4.1	Datamodel	12
4.2	Beregninger og Visualiseringer	13
4.2.1	Energiudnyttelse (kWh/m2)	14
4.2.2	Ferielukning	17
4.2.3	Passivt elforbrug	19
4.2.4	Returtemperatur & Afkøling	19
4.2.5	Anvendelsesgrad	21
4.2.6	Overbliksside	23
5.	Driftsoptimeringsmodellen	24
5.1	Baggrund	24
5.2	Proces	25
5.3	Domæne	25
5.4	Data og associerede begrænsninger	26
5.5	Eksplorativ dataanalyse	27
5.5.1	Dataintegritet	27
5.5.2	Tidsintervaller	29
5.6	Tilgang og designvalg	31
5.6.1	Samlede designvalg	33
6.	Modellering	33
6.1	Valg af algoritme	34

6.2	Isolation Forest.....	34
6.3	Transformationer	35
6.3.1	Heuristisk estimering af anvendelsesgraden	36
6.3.2	Filtrering af ekstreme værdier	37
6.3.3	Interpolation af tidsintervaller	37
6.3.4	Interpoleringsalgoritme.....	38
6.3.5	Filtrering af konstante tidsintervaller	38
6.3.6	Filtrering af inkonsistente dage	39
6.3.7	Tidsgrupperet databerigelse	39
6.3.8	Tidsrelaterede variabler	41
6.3.9	Heuristiske regler	41
6.4	Samlet flow.....	42
6.5	Evaluering og dashboard	43
6.6	Arkitektur	45
6.6.1	Modelregister	46
6.6.2	Integrationer og autentifikation.....	46
6.7	Konklusion.....	47
7.	Energibevidst AI Booking	48
7.1	Baggrund.....	48
7.2	Overblik over EAB	48
7.3	Proces	49
7.4	Data og associerede begrænsninger	49
7.4.1	Lokalets placering	50
7.4.2	Lokalets brugshistorik	50
7.4.3	Lokalets faciliteter	51
7.4.4	Lokalets energiforbrug	51
7.5	Tilgang og designvalg	52
7.5.1	Samlede designvalg.....	53
7.6	Eksperimenter	53
7.6.1	Omkostningsberegning	53
7.6.2	Energiestimering	55
7.6.3	Energi – Naboskab.....	55
7.6.4	Energi – Regression.....	57
7.7	Bookingsystemet.....	58
7.7.1	Kapacitet	59
7.7.2	Faciliteter	59
7.7.3	Anbefalinger	60
7.7.4	Anbefalinger i realtid	61
7.8	Evaluering og perspektiver	61

7.9	Konklusion.....	62
8.	Generelle udfordringer med data	63
9.	Fremtidige perspektiver.....	64
9.1	Analytics	64
9.2	Driftsoptimeringsmodellen.....	64
9.2.1	Evaluering og annoteret data	65
9.2.2	Dataintegritet.....	65
9.2.3	Onboarding af nye bygninger.....	65
9.3	Energibevidst AI Booking	66
9.3.1	Dataintegritet.....	66
9.3.2	Tilpasning og videreudvikling af omkostningsmodeller	66
9.3.3	Optimering af energiberegninger	66
9.3.4	Evaluering	67
10.	Appendix A.....	68

List of Figures

Figur 1: Arkitektur	7
Figur 2: Platformens portal	8
Figur 3: Datamodel til BI	13
Figur 4: Energiudnyttelse (År).....	14
Figur 5: Energiudnyttelse (Dagstype)	15
Figur 6: Energiudnyttelse (Ideal-kurve)	16
Figur 7: Energiudnyttelse (Anvendelsesgrad)	16
Figur 8: Ferielukning (i år mod sidste år).....	17
Figur 9: Ferielukning (hverdag imod weekend)	18
Figur 10: Ferielukning (fast måltal)	18
Figur 11: Passivt elforbrug.....	19
Figur 12: Returtemperatur	20
Figur 13: Afkøling	21
Figur 14: Anvendelsesgrad (bookinger)	22
Figur 15: Anvendelsesgrad (generelt)	22
Figur 16: Overbliksside	24

1. Introduktion til projektet

AI Signaturprojektet har taget udgangspunkt i at undersøge hvordan man med hjælp af AI og datavisualisering kan give kommuner, herunder serviceledere og energikonsulenter, et indblik i og overblik over energiforbruget på tre udvalgte folkeskoler i hhv. Aarhus, Favrskov og Syddjurs kommune.

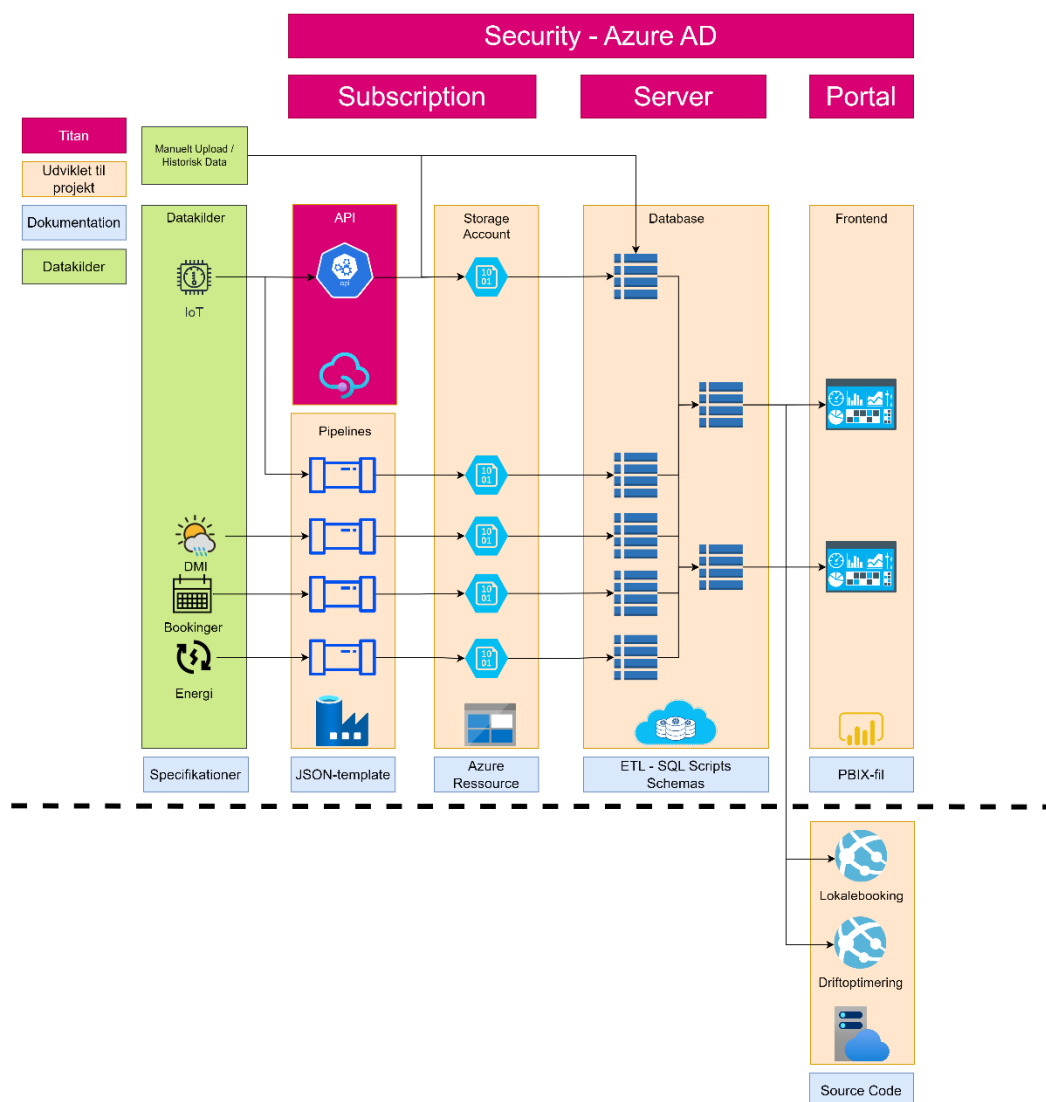
Det har været en eksplorativ tilgang, for i første omgang at undersøge hvilke datakilder kommunerne kunne stille til rådighed for projektet. Herefter har vi tilpasset scopet for leverancerne, således de giver mest mulig værdi i forhold til at give indblik i og overblik over energiforbrug. I de endelige leverancer har slutbrugerens handlekraft været den styrende faktor, hvorfor BI rapporteringen (energisladrehanke) er bygget for at energikonsulenter og skoleledere kan handle på de indsigter de får i leverancerne.

Leverancerne i projektet er:

1. Energisladrehanke Enformanten (Visualisering af data i PowerBI rapport)
2. Driftsoptimeringsmodel (AI model for beregning af anvendelsesgraden for lokaler)
3. Dokumentation for oprindelig scopet AI model (Lokalebooking)

2. Infrastruktur

Projektet er bygget på en hybrid-tilgang, hvor der er taget udgangspunkt i en eksisterende dataplatform udviklet af NTT DATA. Platformen har gjort det muligt at omdanne data til indsigter og værdi hurtigere end hvis man skulle bygge en platform fra bunden til at håndtere data. Samtidigt har platformen givet mulighed for at have et fælles præsentationslag, hvor kommunerne kan tilgå indsigter. I Figur 1 kan ses en overordnet arkitektur for den infrastruktur der har været benyttet i projektet. De følgende afsnit vil gennemgå snitfladerne mellem den eksisterende platform og de funktionaliteter der er opsat og bygget i løbet af projektet.



Figur 1: Arkitektur

2.1 Titan

Generelt set er projektet bygget ved hjælp af **Microsoft Azure**, **Snowflake** og **Power Bi**. Dette gør naturligt at dokumentationen af de infrastrukturelle komponenter hovedsageligt vil bestå af templates og fremgangsmåder for at opsætte de samme ressourcer selv, da de ikke er egenudviklet af NTT DATA.

2.1.1 Azure Subscription / Tenant

Hele projektet er bygget på NTT DATAs **tenant** under platformens **directory**, men på en separeret Azure **Subscription** opsat til GovTech.

2.1.2 API

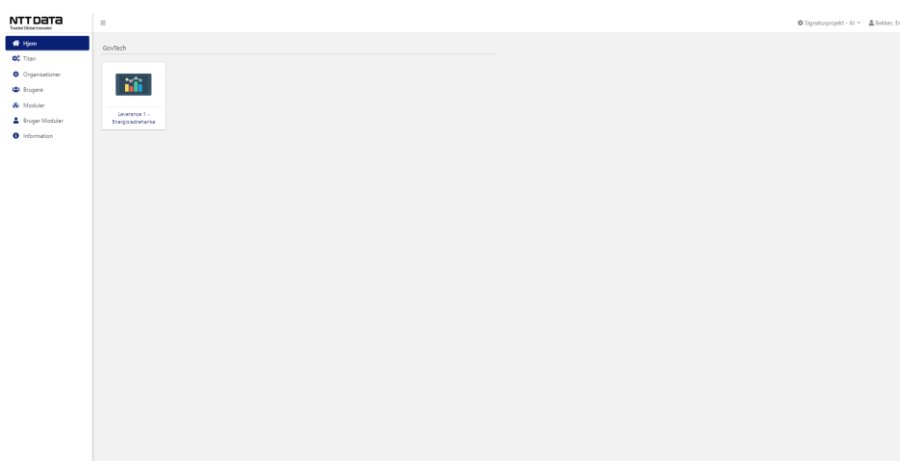
Projektet har haft behov for at benytte et API til at pushe data til platformen fra f.eks. OS2-IoT. APIet der er benyttet er standarden fra den eksisterende platform bygget i **API Management**. Basalt set kaldes APIet via et endpoint og er beskyttet af en API-nøgle. Når der postes en JSON-fil til APIet videresendes det til en Storage Account opsat til projektet. Filnavnene styres dynamisk ud fra tidspunktet for modtagelse af pakken. Der er i projektet dog blevet tilpasset ved at filnavnene styres via tidspunktet i pakken fra Zurface-enhederne, da disse sendes i bulks i intervaller. APIet har været specifikt brugt til at hente IoT-data fra Zurface i Aarhus og Favrskov.

2.1.3 Database

Databasen der er benyttet i projektet, er platformens standard database. Databasen er opsat på platformens instans af **Snowflake** og kan derfor erstattes ved at opsætte sin egen instans på f.eks. Snowflake eller lignende og herpå opsætte samme datagrundlag og transformationer.

2.1.4 Frontend / Portal

Hovedkomponenten som platformen har bidraget med i projektet, er præsentationslaget, som har været selve portalen på platformen. Platformen har her gjort det muligt at logge ind på en portal og tilgå moduler/dashboards. I selve platformen er det også muligt at lave user management og styre hvilke brugere der skal kunne tilgå hvilke moduler. Et eksempel på siden man møder efter login kan ses i Figur 2. Modulerne er opbygget ved hjælp af Embedded Power Bi, som gør det muligt at åbne dashboards alt efter brugeradgang.



Figur 2: Platformens portal

2.2 Udviklet i projektet

Det følgende afsnit vil gennemgå de dele af projektet der infrastrukturelt er blev oprettet/opsat specifikt til projektet og derved vil kunne tages og repliceres.

2.2.1 Dataindsamling / Landing zone

Dataindsamlingen er samlet i en Landing zone som enten indbefatter at data hentes ud af systemer/API'er via data pipelines eller data sendes/pushes til platformen via det før beskrevet API.

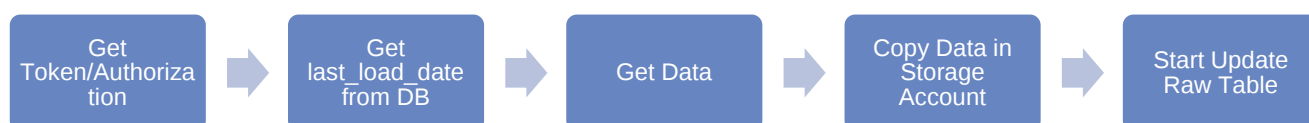
2.2.1.1 Data Pipelines

Kernen af dataindsamlingen er opsat i **Azure Data Factory** som Data Pipelines, der henter og gemmer data i en Storage Account. Der er opsat følgende pipelines til indhentning af data:

Data	Kommune(r)
IoT	Syddjurs
Bookinger	Favrskov, Syddjurs
Vejrdata	Alle
Energidata	Favrskov, Aarhus

Foruden indhentningen af dataene er pipelines også ansvarlig for opdatering af dataene i tabellerne i databasen via ingestion af filerne hentet fra API'er ind i rådata tabeller. Ved de datakilder der selv pusher data til platformens API, f.eks. Zurfice Data fra OS2IoT (Aarhus & Favrskov) bruges pipelines også til at sørge for at disse filer indhentes fra Storage Account til Databasen.

Generelt set er pipelines til indhentning af data opsat efter følgende proces:



Et eksempel på en pipeline, der gør netop dette med Syddjurs' IoT-data:



Når alt data er indhentet, vil pipelines køre en Stored Procedure i Snowflake, som vil persistere de views der er lavet på data der knytter sig til specifikke use cases. F.eks. tages alle bookinger og flades ud på timeniveau i et view, som pipeline vil persistere i en tabel der kan bruges til BI. Persisteringen er med til at optimere performance i BI.

Afslutningsvis bruges pipelines også til at kalde Driftoptimeringsmodellen til både månedlig træning og daglig drift af data efter de er indhentet. Pipelines er opsat til at opdatere hver dag kl. 10.00, men kan konfigureres til f.eks. at være oftere. Datakilderne hentes dog i de intervaller data opsamles i. Der kan i nedenstående ses frekvensen af data fra de enkelte kilder.

Kilde	Aarhus	Syddjurs	Favrskov
Bookinger	Ingen adgang til API	Linje per booking (loades 30 dage frem)	Linje per booking (loades 30 dage frem)
IoT	5-15 min. intervaller	15 min. interval	5-15 min. intervaller
Energi	2 dages forsinkelse på opdatering - Timeniveau	Timeniveau	Timeniveau
Vejr	Timeniveau	Timeniveau	Timeniveau

Alt orkestrering og styring af indhentning og persistering af data styres i en main-pipeline.

2.2.2 Data storage

Som beskrevet i forrige afsnit bruges der to typer af dataopbevaring: **Storage Account** og **Snowflake**.

2.2.2.1 Storage Account

Azure Storage Accounts gør det muligt at opsætte såkaldte Blob Containers, hvori man kosteffektivt kan opbevare semistruktureret filtyper som .csv og .json. Der er i projektet opsat en specifik storage account som har gjort det muligt at separere kommunernes data på platformen.

2.2.3 Frontend – Power Bi

Frontenden der er lavet i projektet, er udviklet i Microsoft Power Bi og er det hovedsagelige visuelle output på dataene. Rapporten er publiceret på platformens tilhørende Power Bi Service som har gjort det muligt integrere løsningen i platformens portal. Selve rapporten ville kunne uploades til en kommunes egen Power Bi Service og derved kunne tilgås derigennem.

2.2.4 AI-modellerne

For at kunne kalde og bruge AI-modellerne (Lokalebooking og Driftoptimering) er disse modeller publiceret til en **Azure Web App**. Dette har gjort det muligt at publicere modellerne i form af Docker Containers og udgive endpoints til at interagere med modellerne. Disse Web Apps er udviklet og opsat specifikt til projektet og er fuldt ud dokumenteret i den tekniske dokumentation

3. Transformationer

Projektet har haft fokus på at tage 3 kommuners data og transformere dem ind i det samme format, således man kan have standardiseret tabeller baseret på typen af data. Dette er gjort for Bookinger, Stamdata (lokaler), Indeklimadata og Energidata. Datoer og vejr er hentet/opsat som fællestabeller for alle kommuner fra starten. Transformationerne fremgår af den tekniske dokumentation og der vil her blot beskrives det endelige dataformat, som der er fundet frem til.

3.1 Fremgangsmåde

Transformationer er opdelt i 2 typer: specifikke og sammenføring. De specifikke transformationer binder sig til de enkelte kommuner og deres lokale datakilder samt formater. Disse transformationer har blandt andet taget metadata fra sensorer og koblet sammen med de enkelte lokaler o.l. Sammenføringen har været kernen i at kunne få alle 3 kommuner ind i et standardiseret dataformat, som kan bruges til en skalerbar udvikling af visualiseringer og analyser baseret på samme datagrundlag mellem kommuner og fremtidigt flere bygninger. De enkelte datakilder i kommunerne har indeholdt forskellige niveauer af information, hvor enkelte kommuner har haft mere data end andre. Der er i sammenføringen lavet tabeller der indeholder de informationer, der går på tværs af kommuner således datagrundlaget bliver fælles.

Databasen er opsat i fire lag som tager dataene fra rå format i "1_RAW" til use case-specifikke tabeller i "4_FEATURIZ". Der er gjort brug af følgende lag:

- 1_RAW
- 2_STANDARD
- 3_CLEANSED
- 4_FEATURIZ

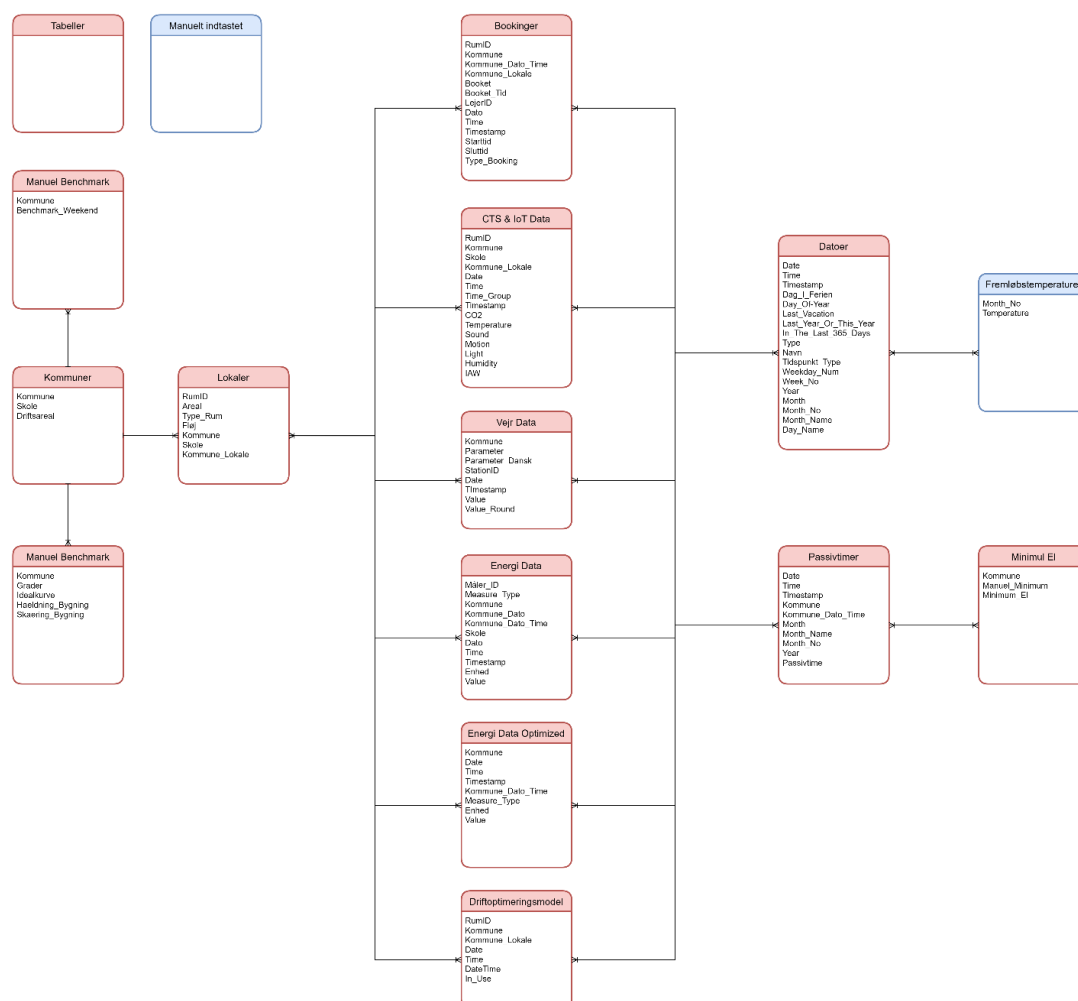
Transformationerne har resulteret i 13 forskellige tabeller, som benyttes til at bygge de dashboards og indsigter der er visualiseret i Power Bi. Datamodellen for disse tabeller fremgår af senere afsnit. De enkelte transformationer fremgår af projektets tekniske dokumentation på GitHub.

4. Business Intelligence/Frontend

Indsigterne og beregningerne er udført og visualiseret i **Microsoft Power Bi**. Det følgende afsnit vil beskrive den opsatte datamodel, som den er sammensat i Power Bi samt de beregninger der er foretaget, for at kunne udregne de udvalgte "energisladrehanke". Afslutningsvis vil siderne gennemgås fra et visuelt og brugsmæssigt perspektiv.

4.1 Datamodel

Datamodellen er baseret på de tabeller der er blevet sammensat i databasen og optimeret til brug i Power Bi. Bookinger, IoT Data, Vejrdata osv. er opsat generisk til at kunne visualisere diverse beregninger og sammenhænge. Performancemæssigt ville man kunne fastlåse sig på et beregnet dataformat som er tilpasset til de historier man gerne vil fortælle i dashboards. Dette er blevet gjort på Passivtimer og Energidata for at vise hvordan man specifikt kan bygge afledte tabeller på sine data, til at kunne besvare de "sladrehanke" der knytter sig til f.eks. el, fjernvarme og passivt forbrug.



Figur 3: Datamodel til BI

4.2 Beregninger og Visualiseringer

Der er i dashboards opsat 5 hovedkategorier også kaldet "sladrehanke", som der analyseres og visualiseres på forskellige niveauer/måder:

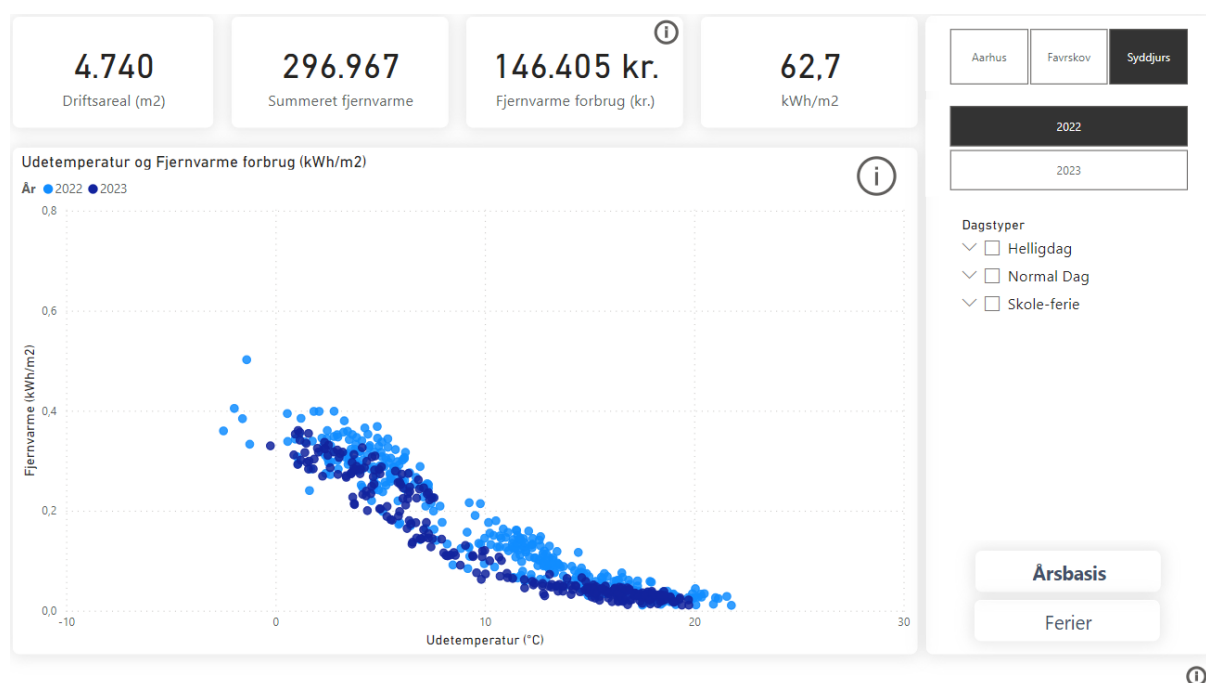
- Energiudnyttelse (kWh/m2)
- Ferielukning
- Passivt elforbrug
- Returtemperatur og afkøling
- Anvendelsesgrad

4.2.1 Energiudnyttelse (kWh/m2)

Energiudnyttelsen er regnet som fjernvarmeforbruget forbrugt i kWh per kvadratmeter. Denne er sammenholdt med temperaturen udenfor, som den er hentet fra DMI på de tætteste vejrstationer for hver skole/kommune. Formlen er således:

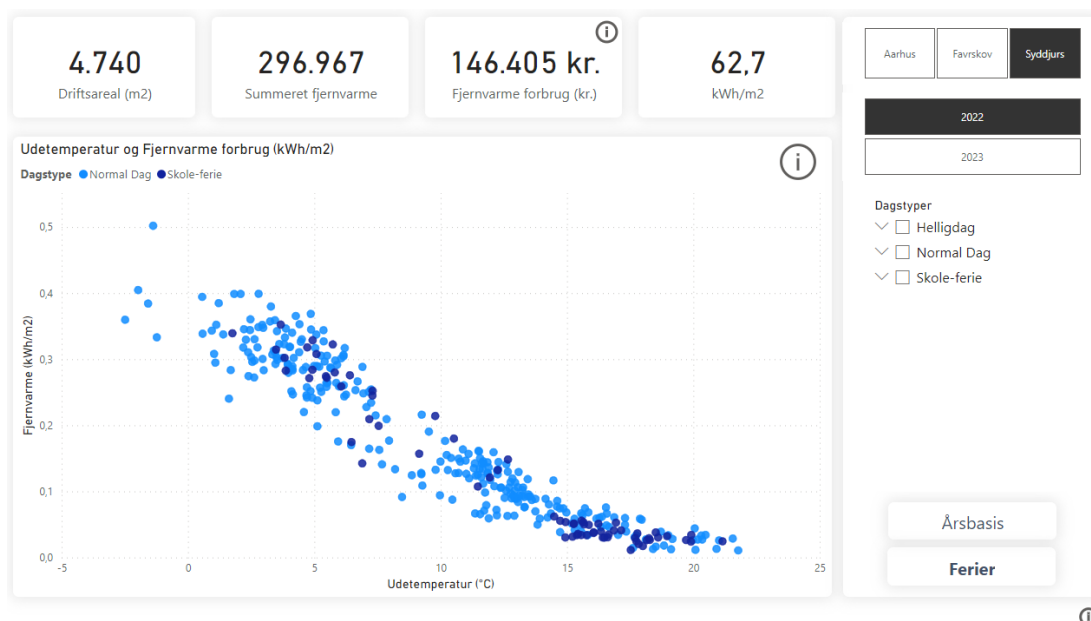
$$\text{Energiudnyttelse} = \frac{\text{Fjernvarme (kWh)}}{\text{Driftsareal (m2)}}$$

Denne er visualiseret på en graf sammenholdt med udetemperaturen. På grafen er det muligt at se 2 år overfor hinanden samt 4 KPI'er på henholdsvis; driftsareal, fjernvarmeforbrug i kWh, fjernvarmeforbrug i kr. og årsgennemsnittet for fjernvarme forbrugt per kvadratmeter. Siden er opstillet som det fremgår af Figur 4.



Figur 4: Energiudnyttelse (År)

Foruden benchmark med året før er energiudnyttelsen også opstillet med farver på om dagene er skoleferier eller normale dage. Denne side fremgår af Figur 5.

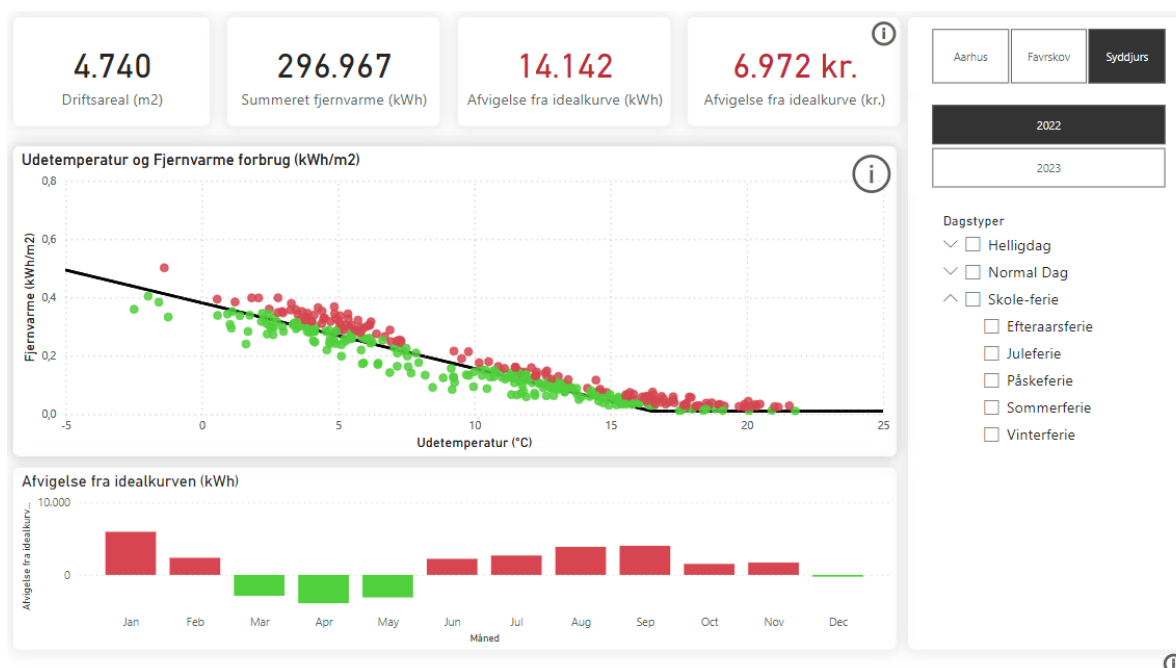


Figur 5: Energiudnyttelse (Dagstype)

Energiudnyttelsen på skolerne er også sammenholdt med en såkaldt "ideal-kurve", som beskriver det optimale fjernvarmeforbrug per kvadratmeter afhængig af udetemperaturen. Ideal-kurven for en bygning er opstillet ud fra følgende formel, hvori t er udetemperaturen, x og b er henholdsvis hældningen og skæringen for den udvalgte bygning:

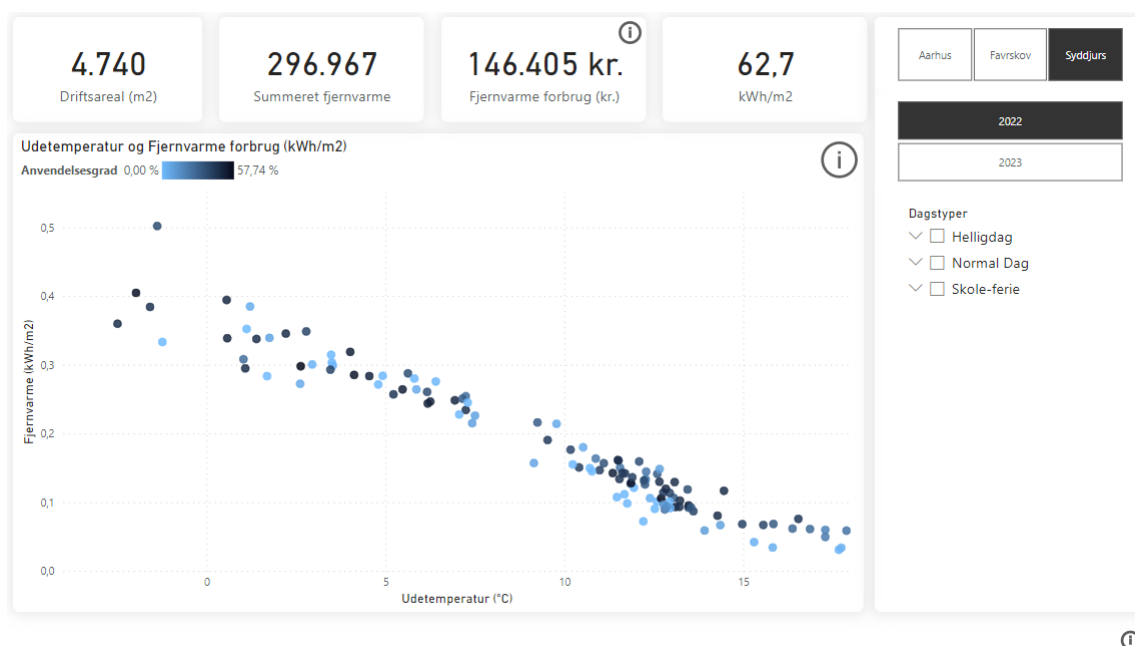
$$\text{Idealkurve} = \text{temperatur} * x + b$$

Denne idealkurve er indtaget på en graf og afvigelsen herfra er udregnet over et år som det fremgår af KPI'en i øverste højre hjørne på Figur 6.



Figur 6: Energiudnyttelse (Ideal-kurve)

Afslutningsvis skal værktøjet bruges til at lave "detektivarbejde" på dage eller perioder hvor man har haft et anormalt fjernvarmeforbrug. Derfor er det nødvendigt at kunne forklare eventuelle udsving, som ikke nødvendigvis er reelle problemer. Dette er forsøgt gjort via indtegning af skolens generelle anvendelsesgrad ud fra Driftoptimeringsmodellen. På Figur 7 ses hvordan en mørkere farve på prikkerne skyldes en forventet højere anvendelsesgrad. Dette gør det muligt at forklare eventuelle udsving ved en højere aktivitet på skolerne.

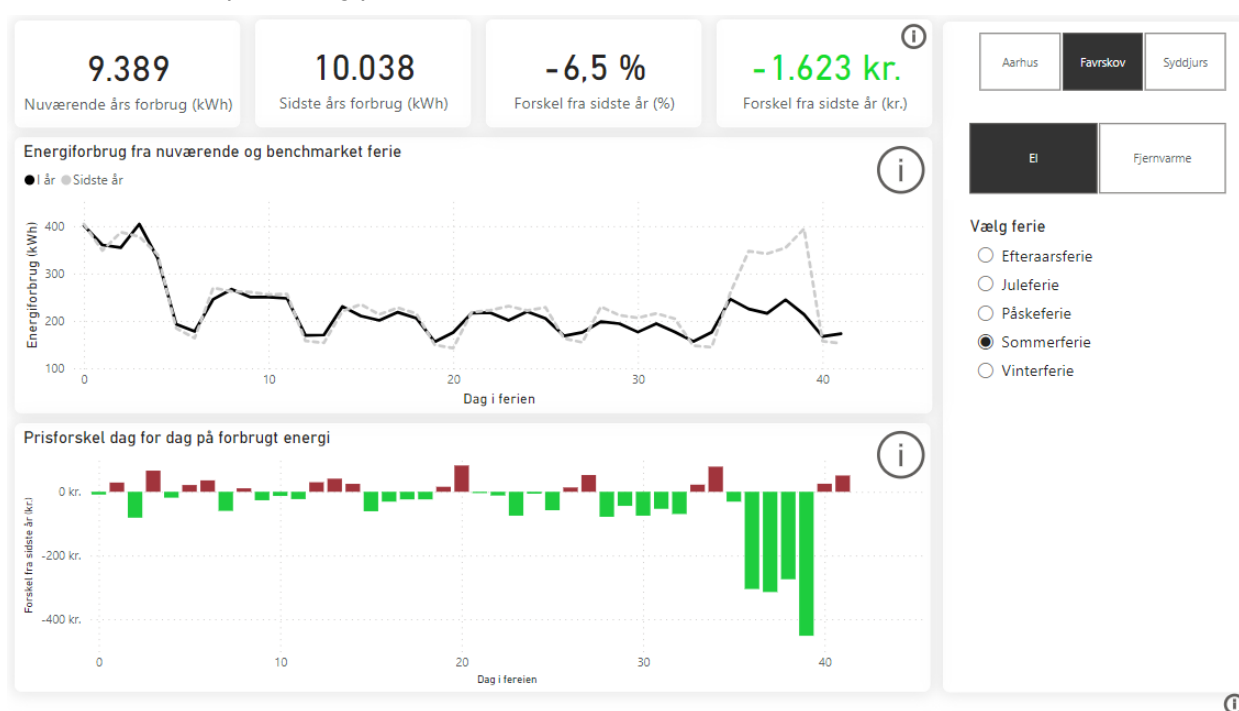


Figur 7: Energiudnyttelse (Anvendelsesgrad)

4.2.2 Ferielukning

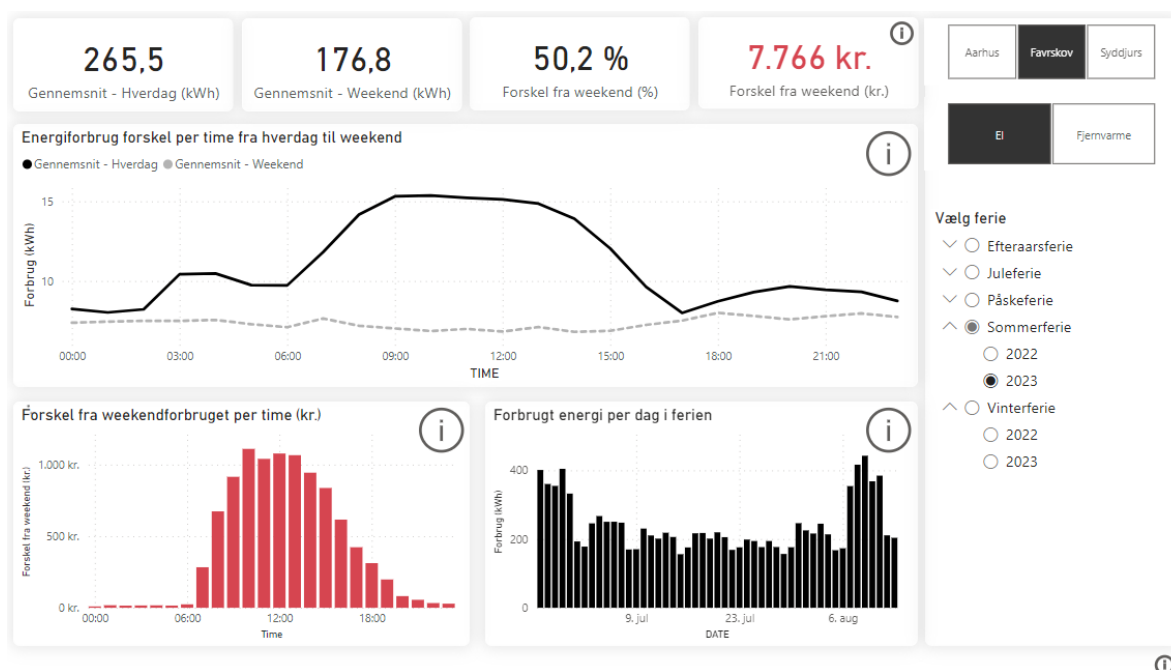
Ferielukningen vurderes ud fra et valgt års feries forbrug på enten fjernvarme eller el som benchmarkes med enten: sidste år, weekender eller et fast måltal.

Den første benchmarking der laves til vurdering af skolernes evne til at optimalt ferielukke ift. energiforbrug er nuværende års forbrug set op imod sidste års forbrug i samme ferie. Der kan generelt på alle sider vælges en kommune, energitype og hvilken ferie der ønskes at analysere. Siden er låst på altid at vise i år overfor sidste år automatisk. Siden viser energiforbruget dag for dag i de 2 år i den øverste graf og differencen i den nederste graf. KPI'erne i toppen af siden viser det samlede forbrug i år og sidste år samt differencen i både procent og pris.



Figur 8: Ferielukning (i år mod sidste år)

Den anden benchmarking der laves, er til at vurdere om skolen ferielukke optimalt ift. at hverdagene gerne skulle ligne weekenderne. Dette skyldes at hverdagene har samme aktivitetsniveau som weekenderne, da der ikke forventes personer på skolen til at udnytte energien. På siden kan der vælges en ferie samt et år for denne ferie. Den øverste graf viser det gennemsnitlige energiforbrug per time i hverdagene imod weekenderne. På grafen i nederste venstre hjørne kan man se hvordan det prismæssigt er forskelligt. Grafen i nederste højre hjørne viser baggrunden for beregningerne, da denne viser den forbrugte energi for hver dag i ferien. KPI'erne i toppen viser, på samme måde som i forrige side, det reelle gennemsnit på de to dagstyper og differencen i procent og kr.



Figur 9: Ferielukning (hverdag imod weekend)

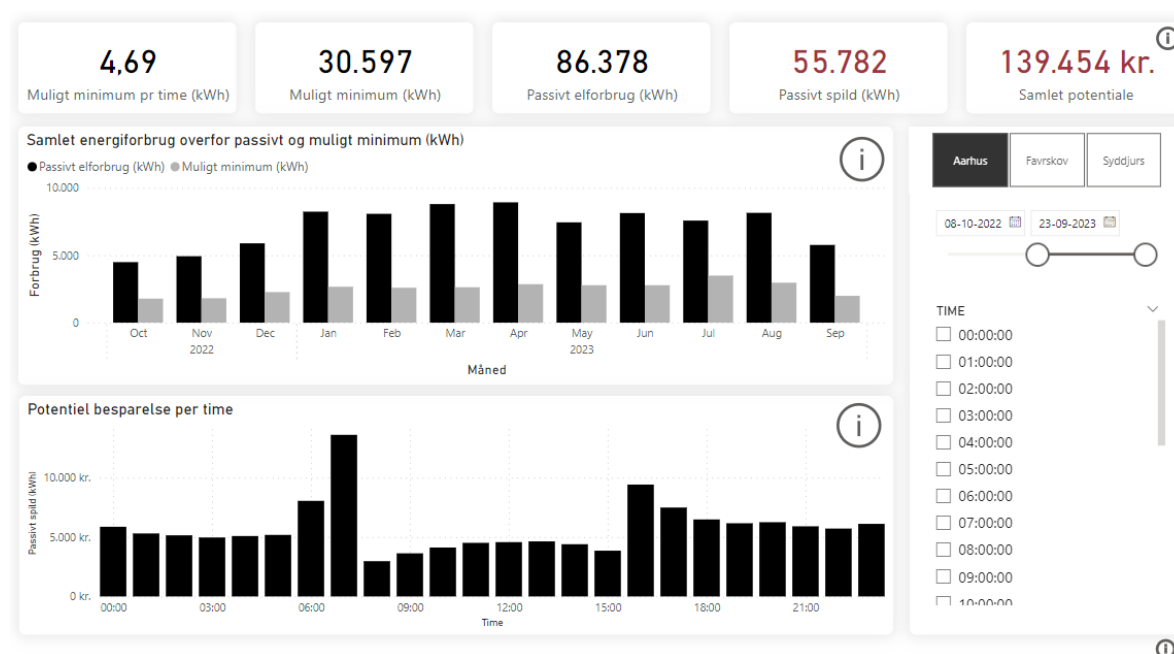
Den sidste benchmark af ferielukningen laves op imod et fastsat måltal fra kommunerne. Her kan man se hverdagene ferien op imod det mål man har fastsat. På grafen øverst kan man se energiforbruget ift. måltallet dag for dag og nederst kan man se det omregnet til en difference i kr. KPI'erne i toppen viser, på samme måde som i forrige side, det reelle gennemsnit på og måltallet samt differencen i procent og kr.



Figur 10: Ferielukning (fast måltal)

4.2.3 Passivt elforbrug

Siden viser en udregning af det passive elforbrug over en valgt tidsperiode. Passivtimer er defineret som timer udenfor skoledage (8-16) hvor der ikke forekommer bookinger af lokaler. Dette gør det muligt at se det mulige beregnede minimum ud fra de 100 timer med lavest elforbrug i nattetimerne overfor det reelle passive elforbrug. Der er stort potentiale i at få skolen ned i passivt elforbrug da denne energi ikke bruges på mennesker eller ophold i bygningerne. Siden viser i den øverste graf måned for måned det reelle passive energiforbrug overfor det beregnede minimum. Det beregnede minimum er som sagt gennemsnittet af de 100 timer med lavest elforbrug i løbet af natten. Dette er ganget med antallet af forventede passivtimer for hver måned. Det beregnede mulige minimum fremgår af de to første KPI'er i toppen. Her kan det også ses hvad reelle passive elforbrug har været samt difference/spildet mellem minimum og reelt forbrug. På den nederste graf er det muligt at se time for time hvilke timer der er de største syndere med hensyn til passivt spild. I øverste højre hjørne er den valgte tidsperiode summeret i kr. på det forventede passive spild/potentiale.



Figur 11: Passivt elforbrug

4.2.4 Returtemperatur & Afkøling

Siderne giver indblik i returtemperaturen og afkølingen af fjernvarme beregnet på en skoles hoved måler. Da der forekommer en afgift for en høj returtemperatur eller for lav afkøling (bonus ved omvendt), er det relevant at se værdierne over året og derved få indblik i den forventede afgift eller tilskud som opnås ved god/dårlig drift. Siderne viser således overordnet den gennemsnitlige værdi for enten returtemperatur eller afkøling for

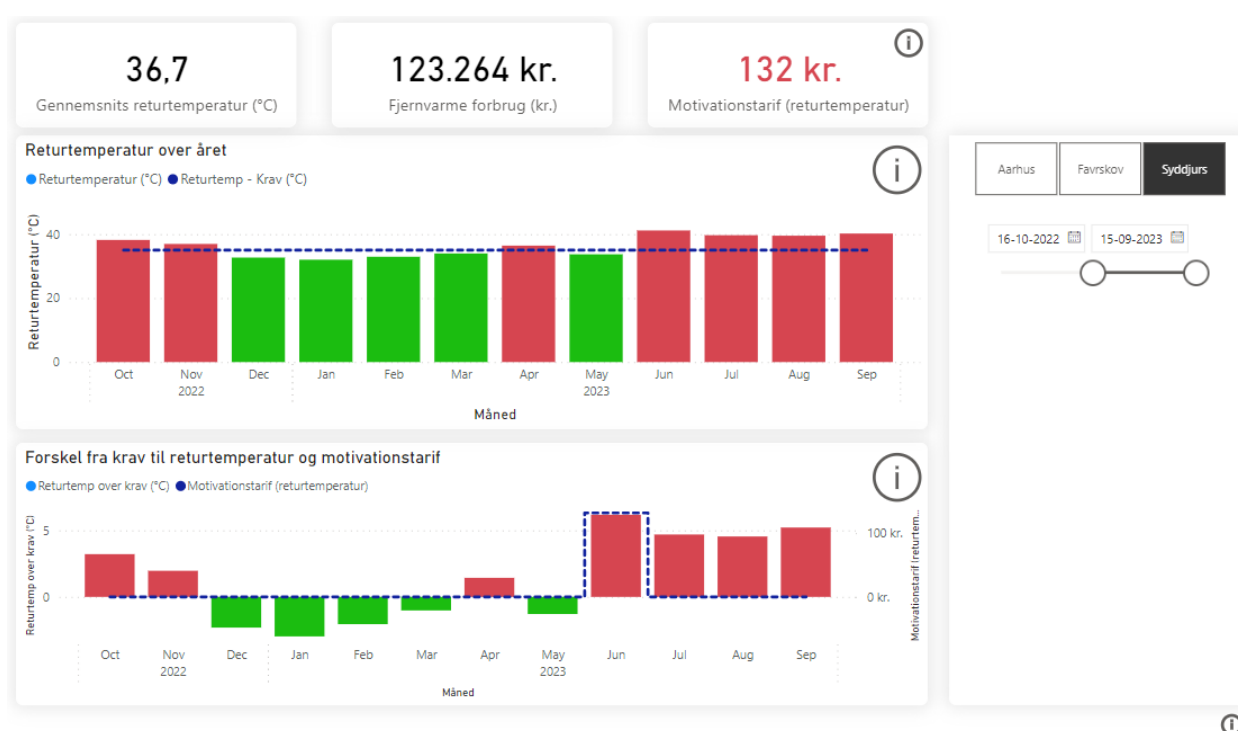
den valgte tidsperiode. Dette er mere granuleret vist måned for måned på den øverste graf.

Straffen/motivationstariffen (mt) er udregnet ud fra følgende formel:

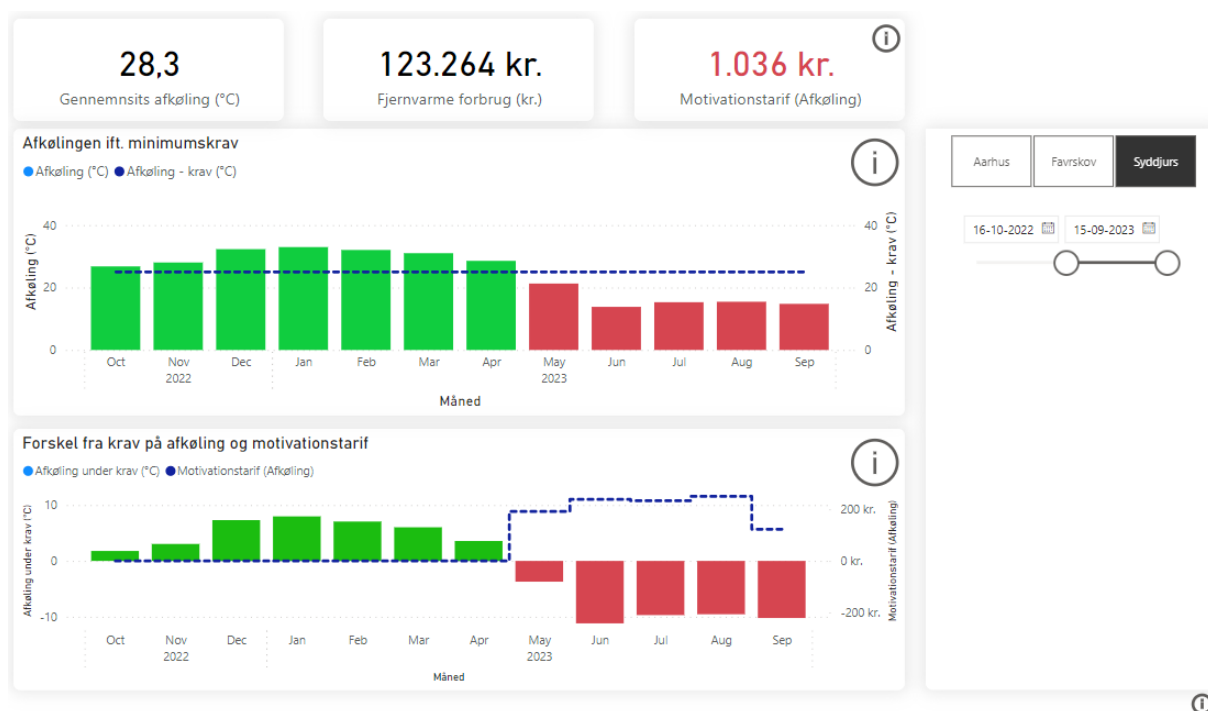
$$mt = t_{\text{målt}} - t_{\text{krav}} * \text{tarifsats (1\%)} * \text{Forbrug (kWh)}$$

Kravene differentierer sig mellem fjernvarmeværkerne. I rapporterne er der lagt en variabel der kan ændres, men ved returtemperaturen er kravet på maksimalt 40 grader for straf og maksimalt 30 grader for bonus. På afkølingen er kravet på minimum 25 grader før man får en straf, som udgangspunkt. Tariffen er udregnet måned for måned og visualiseret i den nederste graf. Summen fremgår af sidste KPI øverst.

De 2 sider til returtemperatur og afkøling fremgår af nedenstående figurer.



Figur 12: Returtemperatur

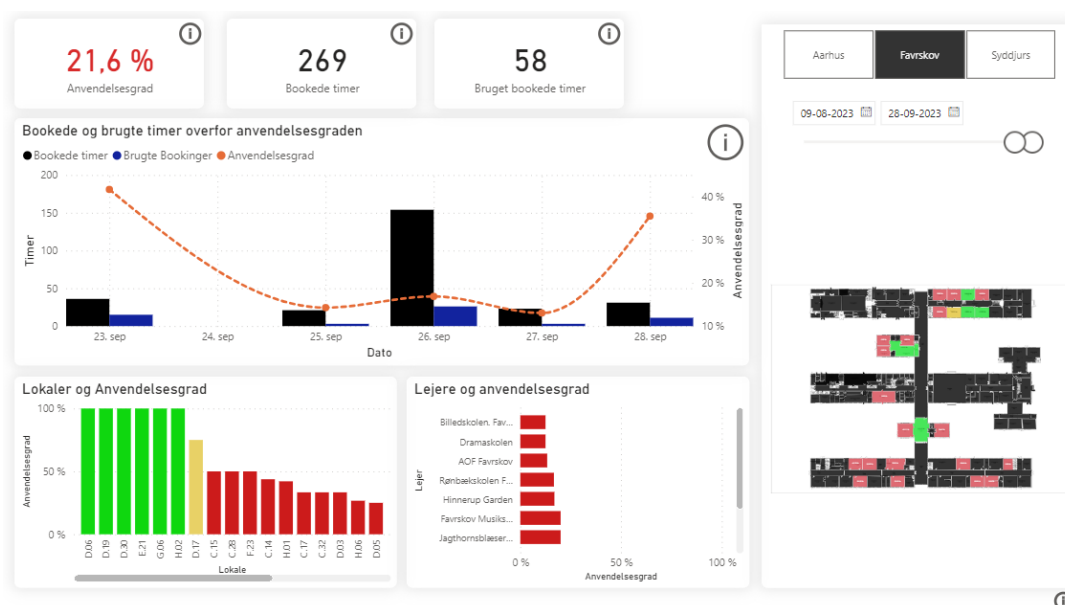


Figur 13: Afkøling

4.2.5 Anvendelsesgrad

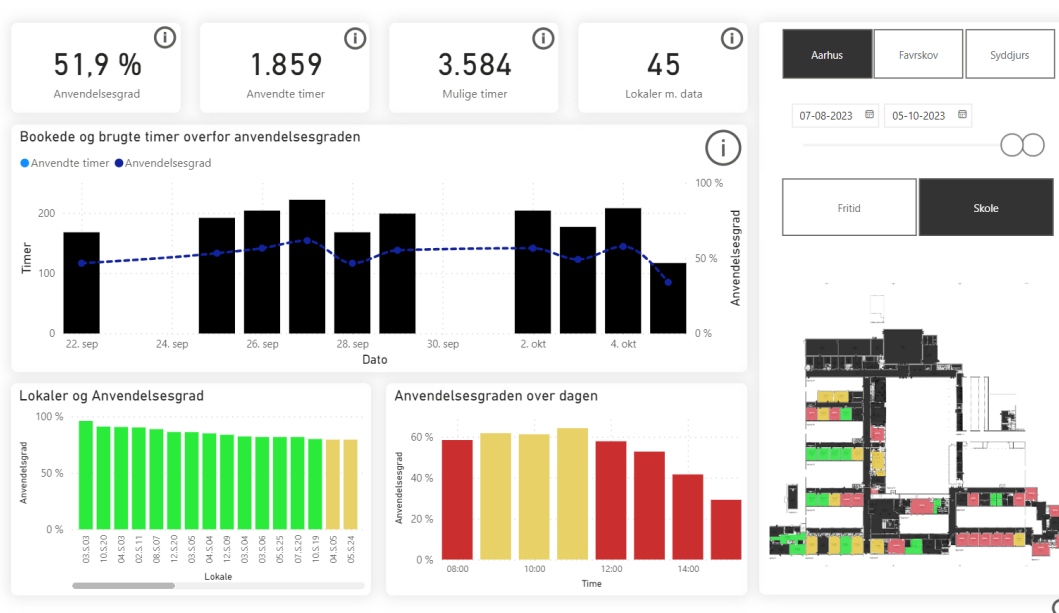
Siderne til anvendelsesgraden er inddelt i 2 men giver generelt giver indblik i skolens forventede anvendelsesgrad ud fra en Driftoptimeringsmodellen. Modellen bruger IoT-sensorer til at definere om et lokale har forventet at være anvendt på baggrund af udvikling i f.eks. CO2-niveauet (se også afsnit om AI-modellerne). Siden gør det muligt at få indsigt i hvor mange timer skolen reelt bliver brugt både i skoletimerne (8-16) og generelt set også i fritiden og ferier samt vurdere om de foretaget bookinger på skolen også er taget i brug.

Den første side knytter sig til netop vurdering af brug af bookinger på skolerne. Her er outputtet af AI-modellen på om lokalet er vurderet i brug eller ej holdt op imod de bookinger som projektet har fået adgang til via f.eks. API'er. Man kan på siden i den øverste graf se for hver dag hvor mange timer der har været booket og hvor mange der har været i brug. Ved at tage forholdet mellem de 2 er der udregnet en anvendelsesgrad i procent. På nederste graf i venstre hjørne kan der ses hvilke bookede lokaler der har de højeste og laveste anvendelsesgrader. Ligeledes kan man ved siden af se hvilke lejere der har størst tendens til at "udeblive". Der kan i toppen ses den gennemsnitlige anvendelsesgrad for den valgte skole/kommune samt summen af de bookede og brugte timer. I højre side kan man se en visualisering af skolen farvet efter anvendelses. Generelt på både denne og næste side betyder farverne følgende: Grøn: over 80% - Gul: 60-80% - Rød: under 60%



Figur 14: Anvendelsesgrad (bookinger)

Den anden side mht. Anvendelsesgrad analyserer helt generelt anvendelsen af alle lokaler med data der har muliggjort en forudsigelse af i brug eller ej af modellen. Her kan man se dag for dag hvor mange timer skolen har været anvendt farvet efter hverdag og weekend. Skolen kan analyseres overordnet i 2 tidsrum: skole og fritid. Siden gør det derfor muligt at få indsigt i hvor mange timer skolen reelt bliver brugt både i skoletimerne (8-16) og generelt set også i fritiden og ferier. På siden kan der også i de nederste grafer ses lokalernes individuelle anvendelsesgrader samt hvilke timer i løbet af en dag lokalene er mest i brug. Øverst kan der ses KPI'er på den gennemsnitlige anvendelsesgrad samt hvor mange timer der har været anvendt, hvor mange timer det har været muligt at forudsige anvendelse samt hvor mange lokaler der har data/sensorer.



Figur 15: Anvendelsesgrad (generelt)

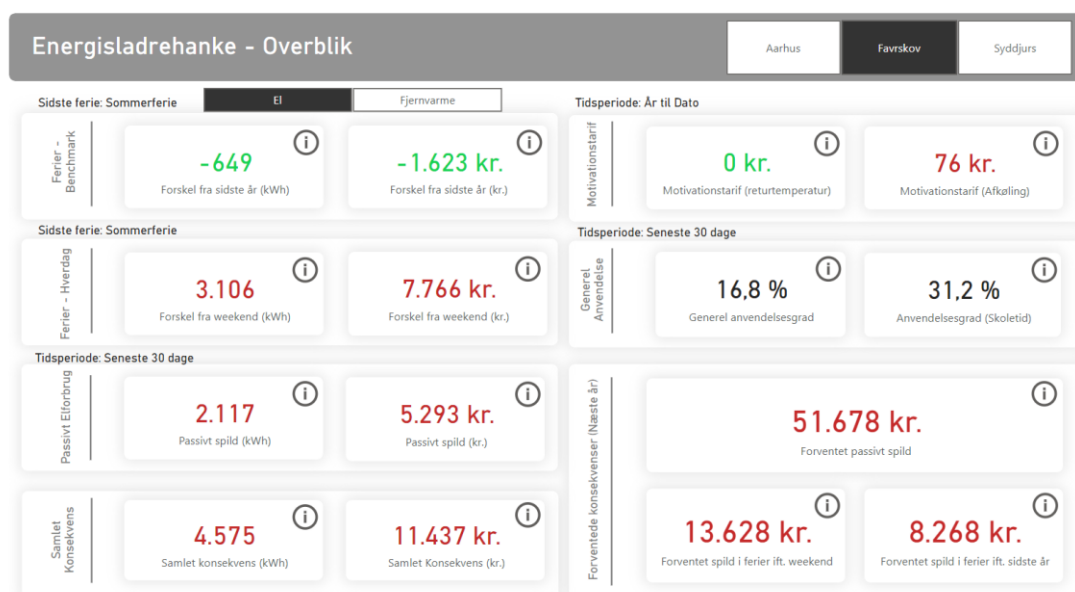
4.2.6 Overbliksside

Overblikssiden er kernen af værktøjet. Her er alle de relevante sladrehanke samlet til ét enkelt overblik, der skal give hurtig indsigt i om man som energikonsulent eller teknisk serviceleder skal reagere. Der er 7 overordnet hovedkategorier på siden.

Den første kategori er benchmark af ferie fra i år til sidste år. Her låser KPI'erne sig automatisk på den seneste ferie og viser forskellen fra året før. Der kan ses forskel både i kWh og kr. (2,5 kr. per kWh). Under dette kan man se den seneste feries forskel mellem hverdag og weekend. Dette er også vist som både kWh og kr. Den tredje gruppe knytter sig til det passive elforbrug. Her kan der ses i kWh og kr. hvor meget skolen forventes at have spildt i de passive timer de seneste 30 dage. Nederst i den venstre kolonne af sladrehanke er de ovenstående grupper regnet sammen til en samlet konsekvens af de seneste 30 dage i passivt spild og den seneste ferie.

Højre side har 2 kategorier af sladrehanke samt en fremskrivning af konsekvenserne. Øverst ses motivationstarifferne på returtemperatur og afkøling summeret til år-til-dato. Under dette kan man se skolens anvendelsesgrad generelt og indenfor skoletid.

Nederst i højre side er konsekvenserne af det passive spild og ferielukning mht. hverdag og weekend samt ift. sidste år forsøgt fremskrevet de næste 12 måneder. Det forventede spild i kr. er de timer hvor skolen er passiv (timer udenfor skoletid uden booking) de næste 12 måneder. Regnet ud fra at antallet af passivtimer det næste år forventes at være det samme som det seneste år. Spildet er beregnet ud fra det gennemsnitlige passive spild per time de sidste 30 dage. Det forventede spild det næste år ift. at hverdagene i ferier forbruger mere energi end weekenderne er beregnet ud fra at ferierne i år vil ligne ferierne næste år, hvis man ikke ændrer noget. Afslutningsvis forventes ferierne i løbet af det sidste år at have samme konsekvens som ferien det næste år hvis intet gøres. I år vil der derfor blive benchmarket med 2022 og det samme vil gøres i de fremtidige konsekvenser når man fremskriver. Disse 3 fremskrivninger fremgår af de 3 KPI'er. Der kan skiftes mellem el og fjernvarme i øverste venstre side, og dette vil påvirke alle udregninger knyttet til de 2 typer forbrug.



Figur 16: Overbliksside

5. Driftoptimeringsmodellen

Den følgende sektion forklarer Driftoptimeringsmodellen (DOM): En AI-applikation til at detektere rumanvendelse, udviklet under AI signaturprojektet *Optimeret Bygningsanvendelse*. Bemærk at brugsdokumentationen for løsningen er at finde under den supplerende tekniske dokumentation.

5.1 Baggrund

DOM blev oprindeligt konceptualiseret og bestilt af de deltagende kommunerepræsentanterne i signaturprojektets arbejdsgruppe, i forbindelse med en generel fokusændring for AI-relaterede aktiviteter. Hvor den tidligere aftalte løsning (AI-bookingsystemet, uddybet nedenfor) var eksplorativ i natur, er visionen bag DOM langt mere konkret: Vi ønsker at anvende sensordata til at detektere hvorvidt et lokale er i brug eller ej, på et givent tidspunkt.

Det er muligt at opstille en simpel regelbaseret model for hvornår et givent lokale siges at være i brug. F.eks. ved at definere en grænse for PPM CO₂, og kategorisere alle tidsintervaller over denne grænse som værende i brug. Udforskes datagrundlaget opdages det dog hurtigt at en regelbaseret model vil komme til kort. For eksempel varierer spændet for CO₂-niveauer betydeligt imellem lokaler: Større lokaler fyldes langsommere, og når måske aldrig denne grænse, imens mindre lokaler fyldes hurtigere men til gengæld har et højt CO₂-niveau lang tid efter det har været i brug. Ligeledes spiller ventilationsanlæg en betydelig rolle: Et velventileret lokale vil have et generelt lavere CO₂-niveau end dets ikke-ventilerede modpart. Formålet med DOM er netop at bygge en løsning der kan genkende robuste mønstre i sensordata, uden at tage

simplificerede valg som, i yderste konsekvens, kan give et fejlbehæftede indblik i anvendelsesgraden på tværs af lokaler.

Praktisk set skal DOM understøtte det bærende *business intelligence*-modul i Enformanten (se ovenfor). Ved at forstå anvendelsesgraden i de respektive lokaler på tværs af skoler og kommuner, kan vi vurdere energiforbruget relativt til kapacitetsudnyttelse. Denne indsigt kan dels være direkte beslutningsstøttende for driftsansvarlige, og dels bruges i afledte analyser og energioptimeringsløsninger (Så som AI-bookingsystemet). Ligesom med AI-bookingsystemet og bi-modulet, søger vi at skabe en løsning der kan gå på tværs af kommuner; en fællesnævner som f.eks. ikke afhænger af datakilder der udelukkende er tilgængelige for en enkelt deltagende kommune.

Som subleverance under AI signaturprojektet *Optimeret Bygningsanvendelse*, skal DOM altså indgå i en større helhed, men behandles i nedenstående som en isoleret løsning.

5.2 Proces

Idet DOM er affødt at et sent fokusskifte i signaturprojektet, har både design- og udviklingsforløbet været relativt kort, sammenlignet med NTT DATAs tidligere udviklingsprojekter af samme kompleksitet. Fordi vi har kunnet tage afsæt i de erfaringer med både domænet og datagrundlaget som vi har gjort os forud for fokusskiftet, har det været muligt at konstruere en AI-baseret løsning til formålet uden en rigid foranalyse og en detaljeret kravspecifikation. Udfordringen – og omkostningen – har i stedet været at modellere domænet retvisende givet begrænsningerne i tilgængelige data, uddybet nedenfor.

Under selve udviklingsforløbet arbejder de til formålet dedikerede ressourcer ud fra en såkaldt A-Z udviklingsstrategi. Det vil sige at en initial fuld løsning bygges med en relativt ineffektiv AI-motor, hvorefter delkomponenterne optimeres for efficiens og performance. Denne strategi sikrer at vi altid er leveringsdygtige, og tillader udviklere at evaluere resultaterne af senere optimering op mod tidligere versioner af løsningen.

5.3 Domæne

DOM er designet og konstrueret på baggrunden af det *scope* (omfang) der defineredes for hele signaturprojektet. Vi præsenteres for tre kommuner, som hver indeholder én skole med et variabelt antal lokaler. For en delmængde af hver skoles lokaler har vi adgang til en multivariat tidsserie af sensordata, så som CO₂-niveau, IAQ, temperatur og aktiveringer af bevægelsessensorer i intervallet. Tidsserien for hvert lokale består af 15-minuters intervaller. I denne domænevirkelighed har vi ligeledes adgang til en mængde metadata for hvert lokale og tidsinterval. For eksempel skoleskemaet, der kan bruges til at se der burde have været undervisning i løbet af et givent tidsinterval, og bookingstatus, der indikerer hvorvidt en booking af

lokalet falder helt eller delvist oveni et givent tidsinterval. Se tabel 17 nedenfor for en fuld oversigt over tilgængelige variabler for hvert tidsinterval.

Samtidig er specifikation af hvornår et lokale er i brug vigtig. Per udsagn fra kommunerepræsentanterne i projektgruppen bør vores løsning ikke identificere f.eks. rengøring eller meget kortvarig brug af et lokale som en reel anvendelse af lokalet. Løsning skal altså identificere hvornår et givent lokale er taget i brug til en *aktivitet*, såsom undervisning, møde eller en lignende fritidsaktivitet.

5.4 Data og associerede begrænsninger

Idet signaturprojektet er et samarbejde imellem de deltagende kommuner om at finde tværgående løsninger, følger det naturligt at vi ønsker at anvende datakilder som er konsistente på tværs af kommunernes datalandskab således at vi kan designe en løsning der kan bruges af alle. Denne præmis begrænser intuitivt de datakilder vi kan designe løsningen på baggrund af.

Hvor én kommune, for eksempel, har bevægelsessensorer i enkelte rum, er dette ikke tilfældet for de resterende kommuner. Ligeledes varierer konsistensen – Dvs. antallet af manglende datapunkter – også betydeligt imellem både skoler og de enkelte rum i en skole. Vigtigste af alt, har vi ikke et pålideligt mål for hvorvidt et lokale rent faktisk har været i brug på et givent tidspunkt, selvom metadata så som skoleskemaet og bookingstatus er delvist pålidelige indikationer. Den største udfordring for AI signaturprojektet som helhed har netop været sammensætningen af de involverede kommuners data, og etableringen af en fælles struktur for projektets data.

Fra Data Factory, via Enformantens datapipelines, modtager DOM en række variabler for hvert tidsinterval (dvs. 15 minutters interval) for hvert lokale. Variabeldefinitionerne er vist i figuren nedenfor.

Variabel	Beskrivelse	datatype
Id	Lokale ID	Heltal
Kommune	Tilhørende kommune	Tekst
Skole	Tilhørende skole	Tekst
Date	Dato for tidsinterval	Tekst
Time	Tidspunkt for tidsinterval	Tekst
Dagsnavn	F.eks. Mandag, tirsdag	Tekst
Tidspunktstype	F.eks. Skoletid eller fritid	Tekst
Skemalagt	Hvorvidt tidsintervallet figurerer i skoleskemaet	Binær
Dagstype	Typen af dag, f.eks. Helligdag, mærkedag, skoleferie	Tekst
Dagskategori	Kalenderdag, f.eks. juleferie eller Kristi Himmelfartsdag	Tekst
CO2	CO2-niveauet (PPM) målt i det givne tidsinterval	Flydende punktværdi
Temperatur	Temperaturen (Celsius) målt i det givne tidsinterval	Flydende punktværdi
Bevægelse	Antallet af unikke bevægelser registreret i det givne tidsinterval	Flydende punktværdi
IAQ	Indoor Air Quality-værdi målt i det givne tidsinterval	Flydende punktværdi
Bookingstatus	Hvorvidt en booking er registreret i tidsintervallet	Binær

Tabel 17: Datavariabler sendt til Driftsoptimeringsmodellen

5.5 Eksplorativ dataanalyse

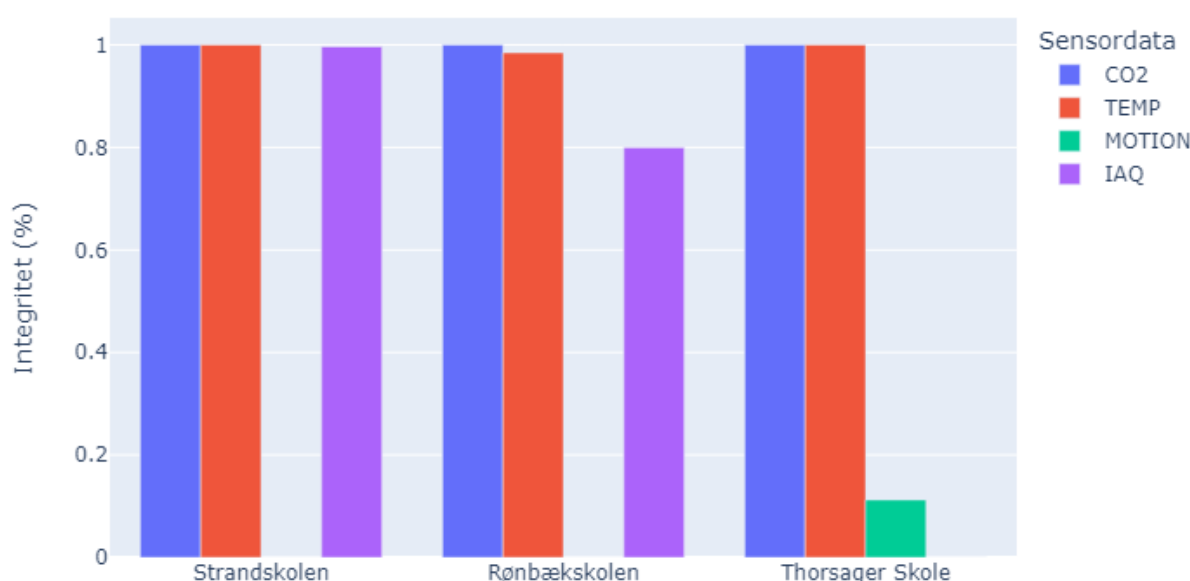
Ser vi bort fra begrænsningerne i de tilgængelige data, har vi effektivt set et domæne med 164 lokaler fordelt på tre skoler, hvor hver skole er tilknyttet én kommune. Som aksiomatisk led i enhver databaseret opgave, udfører vi en eksplorativ dataanalyse for at forstå den tilgængelige data. De følger pointer er plukket fra denne proces, for at give læseren den nødvendige indsigt til at kunne forstå og vurdere de efterfølgende designvalg.

5.5.1 Dataintegritet

Med dataintegritet menes hvor *komplet* den tilgængelige data er, eller hvor mange manglende eller uanvendelige værdier der optræder i en given tidsserie af sensordata. Høj dataintegritet er en præmis for at kunne drage værdi fra data ved hjælp af statistisk modellering. Det følger intuitivt at upålideligt eller manglende data har en negativ indvirkning på en models performance, hvis det ikke direkte umuliggør

modelleringen. De følgende eksempler afdækker dataintegriteten for DOMs datagrundlag, og tjener som baggrund for de senere designvalg foretaget i design- og udviklingsprocessen.

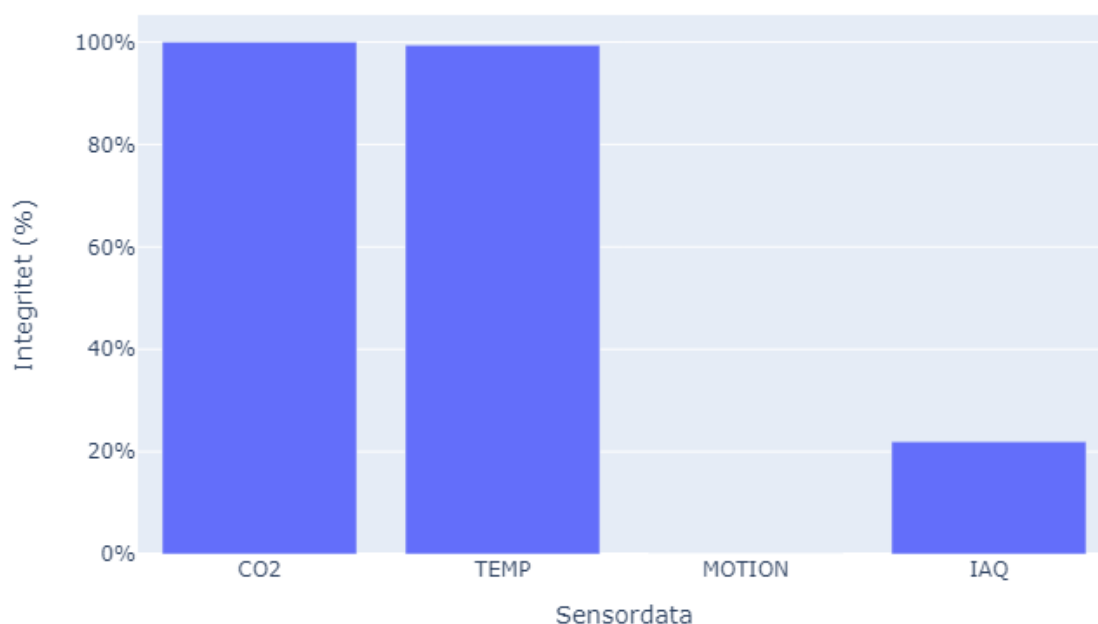
Undersøges dataintegritet på tværs af typer af sensordata per skole (figur 18, nedenfor), ser vi tydelige forskelle imellem typerne. Imens CO2 og temperatur har konsistente høje niveauer af dataintegritet, er virkeligheden en anden for IAQ (Indoor Air Quality) og Motion (Bevægelsessensorer).



Figur 18: Dataintegritet i sensordata per skole

For Strandskolen har vi brugbare IAQ-værdier for over 99% af vores observationer, imens Motion slet ikke figurerer for denne bygning. CO2 og temperatur har nær komplet dataintegritet. For Rønbækskolen har vi ikke Motion-værdier og dataintegritet på omkring 80% og 96% for IAQ og temperatur, imens CO2 er nær komplet. For Thorsager skole besidder vi komplet dataintegritet for både CO2 og temperatur imens IAQ ikke er tilgængelig og Motion har en gennemsnitlig dataintegritet på omkring 10%.

Drejer vi perspektivet til dataintegritet per type af sensordata, uafhængig af hvilke skoler data relaterer sig til, ser vi en lignende tendens. Figur 19 nedenfor visualiserer netop denne indsigt.



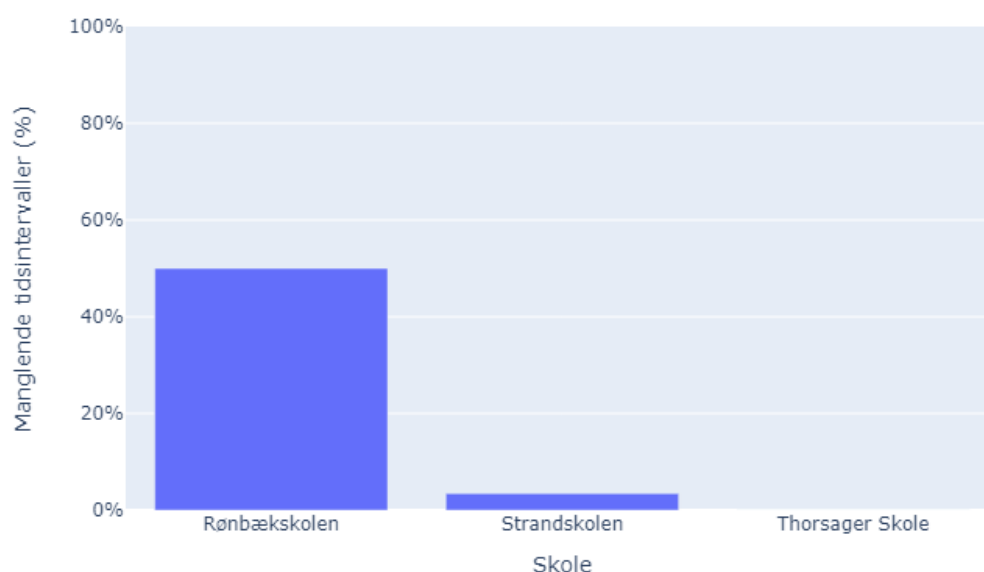
Figur 19: Dataintegritet per type af sensordata

Vi observerer at på tværs af omtrent 1,8 mio. observationer har vi en nær perfekt dataintegritet for CO2 og temperatur imens Motion har brugbare værdier for en negligerbar delmængde af data, og IAQ er relevant for omtrent 22% af de tilgængelige observationer.

Med afsæt i DOMs formål om at skabe en fælles løsning på tværs af de deltagende kommuners datalandskab, reduceres vores råderum til at fokusere på CO2 og temperatur. Dog taler adgangen til Motion og IAQ-værdier for at designe løsningen til at kunne inkludere disse typer data i en senere version hvortil vi kan tilføje disse data med højere dataintegritet.

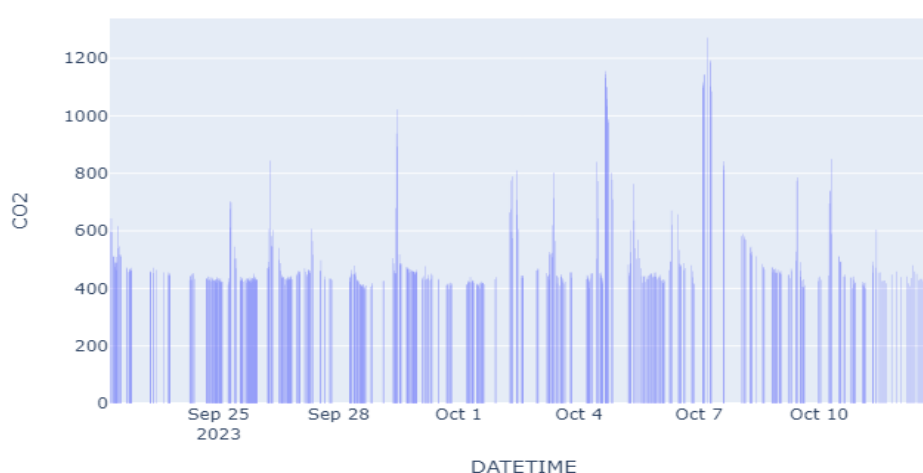
5.5.2 Tidsintervaller

Den tilgængelige data kan betragtes som en kollektion af lokalespecifikke, multidimensionelle tidsserier af sensordata. For hvert lokale har vi n tidsintervaller imellem første og sidste observation i data. Ved nærmere analyse observeres det dog hurtigt at en overvejende majoritet af disse tidsserier indeholder huller – manglende tidsintervaller – og sågar længere perioder af konstante værdier for en eller flere typer af sensordata. Figur 20 nedenfor viser, for hver skole, den gennemsnitlige ratio af manglende tidsintervaller på tværs af lokaler.



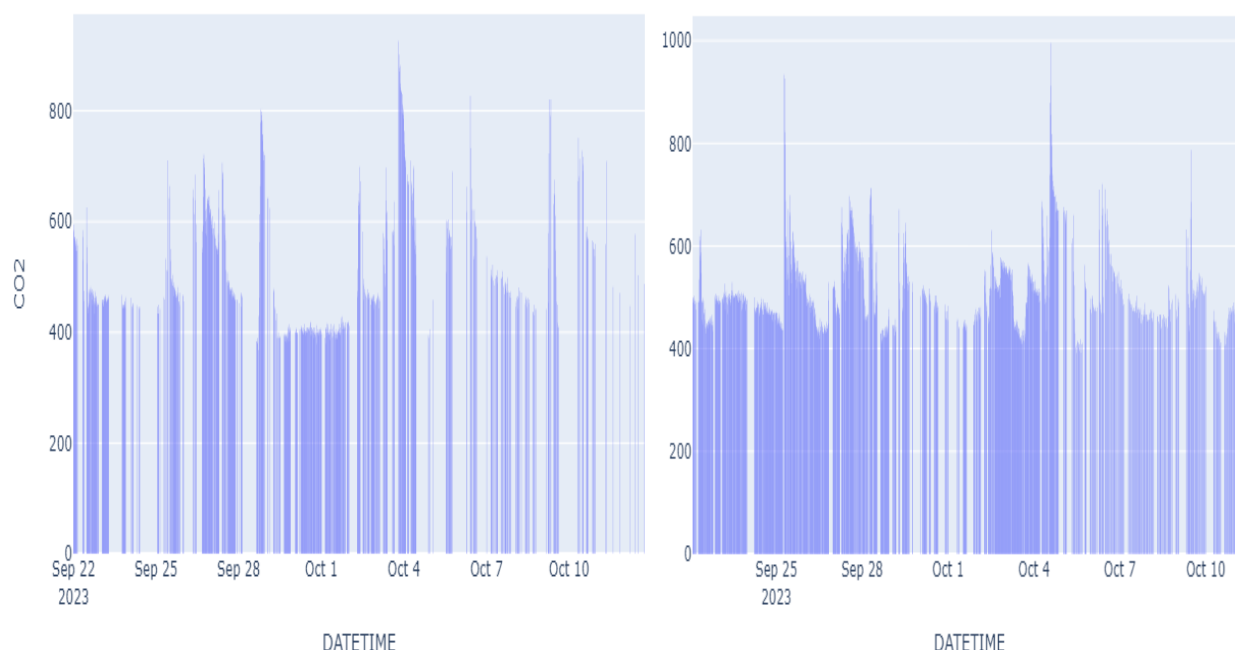
Figur 20: Manglende tidsintervaller (Imellem første og sidste observation)

Vi observerer at lokaler på Rønbækskolen i gennemsnit har lige så mange manglende tidsintervaller som reelle observationer. For Strandskolen og Thorsager skole er denne ratio omkring 2% og mindre end 1%. Imens komplette tidsserier ikke er absolut nødvendigt for modellering, vil større mængder af manglende observationer have en negativ indvirkning på modellernes evne til at fange mønstre på tværs af tidsintervaller. Zoomer vi ind på et enkelt lokale, f.eks. Rønbækskolens lokale E.12 som nedenfor i figur 21, kan vi se hvordan sensordata opfører sig over tid i et konkret eksempel.



Figur 21: Eksempel på manglende tidsintervaller (Lokale E.12, Rønbækskolen) (1)

Lokale E.12 er ikke et enkeltstående tilfælde. Majoriteten af lokaler på samme skole har betydelige huller i tidsserierne af sensordata. To andre eksempler, lokale G.03 og lokale B.26 er givet nedenfor i figur 22.



Figur 22: Eksempler på manglende tidsintervaller (Lokaler G.03 og B.26, Rønbækskolen)

5.6 Tilgang og designvalg

Fra det opstillede domæne og vores eksplorative dataanalyse kan vi identificere fire afgørende faktorer som danner grundlaget for vores tilgang til designet af DOM, opstillet nedenfor, med de primære afledte designvalg for hver faktor.

- **Målet er en indikation af hvorvidt et lokale er i brug eller ej**

For at kunne producere denne type af målværdi på et statistisk grundlag, er vi nødsaget til at tænke over vores løsning som en binær klassificering, til forskel fra f.eks. regressionsalgoritmer der leverer en kontinuerlig målværdi. For at anvende vores historiske data søger vi altså at anvende en statistisk algoritme med et kategorisk output.

- **Ingen adgang til viden om hvorvidt et lokale faktisk var i brug eller ej**

Mens vi har de ovenfor nævnte tidsserier af sensordata, har vi ikke en pålidelig kilde til at fortælle os hvorvidt et lokale rent faktisk var i brug eller ej på indenfor et givent tidsinterval. Selvom både skoleskemaet og bookingstatus intuitivt burde kunne bruges som supporterende kilder til at afdække dette, viser vores analyse en signifikant forskel imellem mønstre i sensordata der indikerer at lokalet har været i brug og disse metadata. Vi har altså ikke pålidelig annoteret data, hvilket gør superviseret maskinlæring komplekst hvis

ikke decideret umuligt. Herfra har vi to muligheder: Vi kan enten investere en substantiel mængde tid til at annotere data for at muliggøre superviseret maskinlæring, eller søge mod superviseret maskinlæring, såsom klyngeanalyse, visse typer af tidsserieanalyse og anormalitetsdetektion, der ikke kræver annoteret data. Grundet den relativt korte tidshorisont for subleverancen har vi valgt at arbejde fremad med usuperviseret maskinlæring. Denne gren af AI præsenterer sine egne udfordringer, så som kompleksitet i evaluering af resultater, hvilke gennemgås nedenfor.

- ***Tilgængelige data er tidsserier af lokale-specifikke, multivariate sensordata***

Tidsseriedata er en speciel kategori af tabulær data, hvor hver observation er knyttet til et specifikt tidspunkt, hvilket gør tiden til en central dimension i datasættet. I modsætning til almindelig tabulær data, hvor rækkefølgen af rækker ikke har stor betydning, er sekvensen i tidsseriedata afgørende, da hvert datapunkt er forbundet til et tidsstempel og ofte afhænger af tidligere observationer. Denne tidsmæssige sekvens skaber en naturlig ordning, som er væsentlig for analyser og modeldannelse. Samtidig er autokorrelation er et fremtrædende træk ved tidsseriedata, som indikerer at værdier tæt på hinanden i tid er mere sandsynlige for at være korreleret end dem, der er langt fra hinanden. Dette adskiller sig fra almen tabulær data, hvori hvert eksempel ofte antages at være uafhængig af de andre. I vores domæne er det netop tendenser i sensordata på tværs af tidsintervaller der indikerer om et lokale er i brug eller ej.

Endvidere er tidsseriedata ofte præget af sæsonmæssige mønstre og tendenser, hvor data viser regelmæssige og forudsigelige mønstre over bestemte tidsperioder. Disse mønstre kan være daglige, ugentlige, månedlige eller endda årlige. Tidsseriedata kan også indeholde langsigtede skift eller cykliske mønstre, der ikke nødvendigvis er bundet til en sæson, men snarere viser op- og nedture over længere tidsintervaller. I vores domæne observerer vi f.eks. cyklicitet i anvendelsen af et lokale i løbet af en uge, hvor mandag til fredag har høj anvendelse mens anvendelse i weekender varierer. Ligeledes forholder det sig med skoleferier og helligdage. Ydermere, adskiller disse mønstre sig imellem lokaler, idet hvert lokale har dels et unikt brugsmønstre og dels unikke attributter, så som størrelse, tilstedeværelse af airconditionlæg og antallet af vinduer. Dette påvirker både de nominelle værdier registreret for hver type sensordata såvel som den relative autokorrelation i hastighed og acceleration med hvilke disse ændrer sig. Derfor søger vi at modellere hvert rum separat, for bedre at kunne afdække lokalet specifikke mønstre.

Per vores scope, ønsker vi samtidig at instruere vores model(ler) til at ignorere antaget anvendelse når der ikke er tale om en reel aktivitet, så som undervisning, møder eller fritidsaktiviteter. Dermed bør f.eks. rengøring eller gennemgang – kortvarig anvendelse – ikke betragtes som anvendelse. For at opnå denne tilstand er vi altså nødsaget til at differentiere imellem kortvarig anvendelse og reelle aktiviteter. I tidsseriemodellering gøres dette typisk ved hjælp af *udglatning* (smoothing): En process igennem hvilken vi udglatter små udsving på tværs af tidsintervaller i hver tidsserie af sensordata og forstærker mønstrene i større udsving på tværs af tidsintervaller.

- ***Tilgængelige sensordata har inkonsistent integritet.***

Udfordringer såsom manglende data eller ekstreme værdier (outliers) kan have betydelig indflydelse på analyser af tidsseriedata og kræver ofte særlige teknikker for at imødegå disse problemer, såsom interpolering eller udeladelse af ekstreme værdier. I vores domæne ser vi f.eks. en vis mængde ekstreme værdier, såsom CO2-niveauer over 8000 PPM eller temperaturer under -40 grader celsius. Disse værdier er åbenlyse sensorfejl som vi må tage højde for. Ligeledes indeholder den tilgængelige data en række manglende tidsintervaller og sensorværdier, som må behandles for at kunne skabe en datarepræsentation af domænet der kan modelleres.

- ***Tilgængelige data varierer på tværs af kommuner og skoler***

Som nævnt ovenfor, er det er præmis for hele AI signaturprojektet at vi tilstræber at konstruere fælles løsninger på tværs af datalandskaber i de involverede kommuner. Fordi mængden, integriteten og typerne af tilgængelige data fra hver kommune varierer, søger vi i DOM at gøre brug af typer af sensordata der er konsistente på tværs af kommunerne. Dette indskrænker naturligvis vores råderum for inddragelse af interessante datakilder så som signaler fra bevægelsessensorer, som desværre kun er tilgængelige for en delmængde af skolerne.

5.6.1 Samlede designvalg

Samlet set søger vi altså at opstille en *binær klassificering af anvendelse af hvert unikt lokale*, ved hjælp af *usuperviseret maskinlæring på multivariat tidsseriedata med høj sæsonbetingelse og cyklicitet*. Den anvendte data skal være *tilgængelig for alle involverede kommuner/skoler*, og skal transformeres og/eller filtreres for at behandle *ekstreme værdier, manglende tidsintervaller* og for at *udglatte mindre udsving der indikerer kortvarigt brug*.

6. Modellering

På baggrund af de samlede designvalg gennemgået ovenfor, har vi nu et klart modeldomæne at designe DOM ud fra. Dette afsnit gennemgår den praktiske operationalisering af ovenstående og de overvejelser der lægger bag disse valg. Et visuelt overblik over det resulterende *flow*, dvs. løsningens process, inklusive transformationer, vises senere i afsnittet "*Samlet flow*" i figur 30.

6.1 Valg af algoritme

Indenfor rammen af usuperviseret maskinlæring til tidsseriemodellering, har vi en bred vifte af algoritmer at eksperimentere med. Den fundamentale intuition bag denne klasse af algoritmer er at forholdet imellem datapunkter kan anvendes til automatisk at opstille diskriminerende beslutningsgrænser, til forskel for superviseret læring som anvender eksisterende eksempler på korrekt klassificerede instanser til at afdække disse beslutningsgrænser. Hvis vi betragter tidsintervaller hvor lokaler ikke er i brug som klasse 0 og tidsintervaller hvor lokaler er i brug som klasse 1, har vi en binær repræsentation af vores domæne. Herfra er valget af den specifikke matematiske algoritme baseret på både ovenstående designvalg og følgende praktiske overvejelser. DOM bør

- Kunne fungere i realtid,
- Være hurtig, simpel og nem at forklare, og
- Robust overfor støj i data.

Anomalitetsdetektion er en prominent klasse af algoritmer under usuperviseret maskinlæring, som har vist sig at være særlig effektive i blandt andet detektion af anvendelse af rum baseret på sensoriske data, som den vi ønsker at udvikle. Anomalitetsdetektion kan netop konfigureres til at fungere i realtid, hvilket er afgørende for at overvåge systemer løbende, og håndtere støj i data hvilket meget vel kan øge nøjagtigheden af vores resulterende indsigter givet den svingende integritet i den tilgængelige data. Algoritmer til anomalitetsdetektion, så som Isolation Forest, er desuden robuste over for variationer i data, byggede til at håndtere højdimensionel data, og kan nemt skaleres, hvilket er afgørende i domæner med store datamængder som vores.

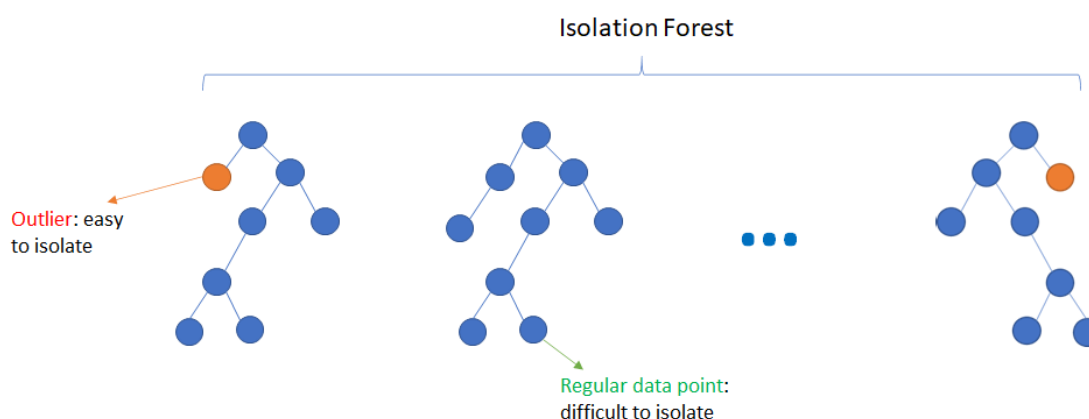
Ved at modellere domænet som en anomalitetsdetektion, kan vi skabe en løsning der identificerer tidsintervaller med usædvanlig aktivitet i sensordata som værende i brug, uden at blive forstyrret af kortvarig eller ikke-signifikant aktivitet, såsom rengøring eller passage gennem rummet.

I lyset af disse overvejelser har vi valgt anomalitetsdetektion, og især Isolation Forest-algoritmen, som kernen i DOM. Denne tilgang tillader os at opstille en række statistiske modeller for identifikation af faktisk rumbrug, hvilket er afgørende for at kunne optimere driftsplanlægningen og ressourceallokeringen i bygninger og på tværs af deltagende kommuner.

6.2 Isolation Forest

Isolation Forest-algoritmen er baseret på princippet om, at anomaliteter er få og forskellige i deres egenskaber sammenlignet med "*normale*" observationer, som f.eks. nattetimer. Et fundamentalt princip i denne algoritme er afhængigheden af et heuristisk estimat af ratioen af anomaliteter i data. Til dette formål anvendes metadata som skoleskema og bookingstatus, som, skønt de ikke er pålidelige nok til at agere

reelle målværdier (annoteret data), kan bruges som input til at opstille et estimat. Denne process beskrives i først subproces under *Transformationer* i næste afsnit. Algoritmen er matematisk beskrevet i den tekniske dokumentation, vist herunder i figur 23, og beskrevet i punktform i det efterfølgende.



Figur 23: Isolation Forest: Identificering af anormaliteter/Ekstreme værdier

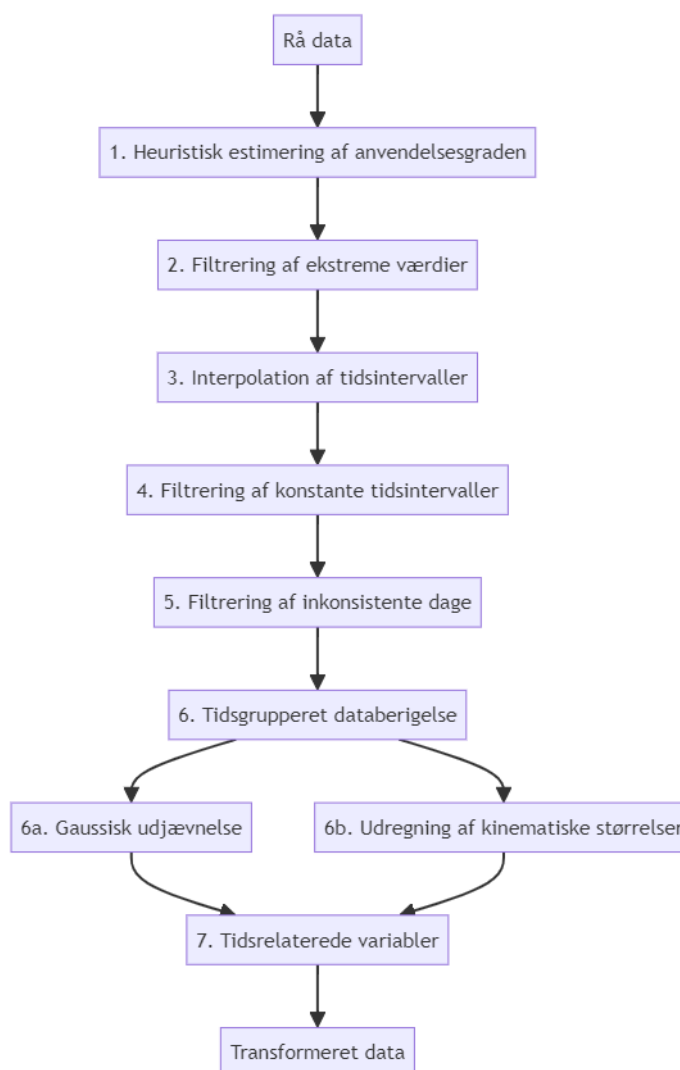
De følgende trin beskriver Isolation Forest-algoritmens overordnede processer.

- **Isolation:** Algoritmen isolerer observationer ved at vælge en egenskab (feature, dvs. en sensorværdi) tilfældigt og derefter tilfældigt vælge en opdelingsværdi mellem de maksimale og minimale værdier af den valgte egenskab.
- **Træ:** Herefter deles data ved ud fra ovenstående opdeling. Denne proces gentages rekursivt, indtil hver observation er isoleret, hvilket skaber et såkaldt "træ" af opdelinger
- **Skov:** Flere sådanne træer bliver kombineret til at danne en "skov". Jo hurtigere en observation bliver isoleret (dvs., jo færre opdelinger der kræves), desto mere sandsynligt er det, at observationen er en anormalitet.
- **Anormalitetsscore:** Hver observation får en score i intervallet $[0, 1]$, der reflekterer dens grad af anormalitet. En høj score (over 0,5) indikerer en stærk anormalitet, mens en lav score indikerer normalitet.

6.3 Transformationer

Med vores valg af algoritme, kan vi konstruere de transformationer som den relevante data skal gennemgå for at kunne anvendes som træningsdata til vores model(ler). Transformationer kan deles op i to kategorier: Rensning og berigelse. Hver subproces i disse kategorier gennemgås nedenfor, hvor 1 – 5 er datarensning og 6 – 7 er databerigelse. Til formålet for denne rapport udelades dybdegående beskrivelser af det eksplorative, iterative arbejde der har resulteret i nedenstående subprocesser. Her refereres til A-Z udviklingsstrategien (forklaret ovenfor) hvori en fuld - om end ineffektiv - løsning bygges først, hvorefter subkomponenterne optimeres for efficiens og performance, og evalueres i sammenligning med tidligere

versioner af løsningen. Det er netop denne strategi der har dirigeret arbejdet imod nedenstående. Nedenstående transformationer er opstillet i den rækkefølge transformationerne finder sted. Den samlede transformationsproces kan repræsenteres som et samlet flow, vist nedenfor i figur 24.



Figur 24: Datatransformationer i sekvens

6.3.1 Heuristisk estimering af anvendelsesgraden

Som nævnt ovenfor kræver Isolation Forest-algoritmen et heuristisk estimat af ratioen af anormaliteter, eller i vores tilfælde antallet af tidsintervaller hvori et givent lokale antages at være i brug. Til dette formål anvender vi to typer af metadata: Skoleskemaet og bookingstatus. Vores estimerende anvendelse for et givent lokale er givet ved

Estimeret anvendelse = $\frac{\text{(Antallet af tidsintervaller hvori der enten er registreret en booking ELLER som optræder i skoleskemaet)}}{\text{Det total antal tidsintervaller i tilgængelige data for lokalet}}$

Husk fra tabel 17 (Variabeldefinitioner) at vi har to binære kategorier tilknyttet hvert tidsinterval for hvert lokale: *Bookingstatus* fortæller os lokalet er booking i netop dette tidsinterval og *Skoleskema* fortæller os om netop dette tidsinterval er en del af de timer der er planlagt til undervisning i skoleskemaet. Ved at undersøge om én af disse to værdier er positive (dvs. sande), har vi et estimat af anvendelsen for det givne lokale. Ved at dividere denne værdi med det totale antal tidsintervaller har vi en normaliseret ratio af anvendelse, eller den antagne anvendelsesgrad for lokalet

6.3.2 Filtrering af ekstreme værdier.

Ekstreme værdier (Så kaldte *outliers*), defineret som datapunkter, der falder uden for forudbestemte grænser opstillet ved hjælp af vores viden om domænet, fjernes fra datasættet. Det er for eksempel alle tidsintervaller hvor CO2-niveauet (i PPM) er under 1 eller højere end 8000. Intuitivt kan vi antage at værdier udenfor disse grænser skyldes sensorfejl. Dette er et datarensningstrin, der sikrer, at modellen ikke bliver skævvredet af ekstreme værdier.

6.3.3 Interpolation af tidsintervaller

Givet den svingende dataintegritet, har vi tre overordnede (og ikke gensidigt udelukkende) muligheder: At fremskaffe data af højere integritet, at fjerne al data af lav integritet eller at interpolere manglende data. Givet den korte udviklingsramme for DOM, er den første mulighed ikke mulig, selvom den er fortrinsvis bedre i et langsigtet perspektiv, som uddybet i et senere afsnit. Den anden mulighed vil eliminere vores evne til at bygge DOM, idet vi ikke besidder en datakilde uden manglende data. Den sidste mulighed, interpolering af manglende data, er derfor vores strategi.

Udfordringer ved interpolering af data i tidsserier inkluderer risikoen for at indføre bias eller fejl, hvis de manglende data er ikke-tilfældigt fordelt eller reflekterer skjulte mønstre. Det kan også fordreje den underliggende tidsafhængige struktur, såsom sæsonmæssighed og trends, og give en falsk følelse af præcision, især hvis perioder eller punkter i den tilgængelige data, hvor information mangler, er omfattende eller systematiske, som i vores tilfælde.

For at undgå at konstruere antifaktiske mønstre i data, begrænser vi mængden af data vi interpolerer. Konkret interpolerer vi op til 3 sammenhængende manglende tidsintervaller. Nedenstående figur viser tre hypotetiske eksempler, hvor NaN (Not a Number) repræsenterer manglende værdier.

450		450	
NaN		477	
NaN	->	530	Interpolerer begge
588		588	sammenhængende værdier
580		580	
[...]		[...]	
450		450	
NaN		477	
NaN	->	530	Interpolere op til tre
NaN		588	sammenhængende værdier
NaN		NaN	
[..]		[..]	
NaN		NaN	
NaN	->	NaN	Interpolerer kun sidst værdier,
NaN		NaN	idet startværdien ikke giver et
NaN		830	grundlag at starte ud fra.
850		850	

Figur 25: Eksempler på interpolation

6.3.4 Interpoleringsalgoritme

For at sikre at vores interpolerede værdier repræsenterer domænet så godt som muligt, anvender vi kubisk kurvet interpolation (*Cubic spline interpolation*) med en tovejsvægtning. Denne algoritme kigger i begge retninger efter datapunkter at basere interpoleringen på, og opstiller en polynomisk kurve imellem to ægte datapunkter på hver side for at danne den syntetiske værdi. Eksisterer der ikke to ægte datapunkter på hver side, falder vores transformation tilbage til lineær interpolation, ligeledes baseret på en tovejsvægtning.

6.3.5 Filtrering af konstante tidsintervaller

Den tilgængelige data indeholder en signifikant mængde af *konstante tidsintervaller*, det vil sige sammenhængende tidsintervaller med en konstant værdi for en eller flere tidsserier af sensordata. Indenfor sensordata, er konstante værdier meget sjældent retvisende. F.eks. er det meget usandsynligt at et givent lokale vil have en eksakt samme PPM CO₂-værdi på tværs af f.eks. tre tidsintervaller (45 minutter), siden CO₂, ligesom de andre typer sensordata, naturligt fluktuerer ved blot den mindste aktivitet i eller omkring lokalet. F.eks. vil en bil på vejen udenfor producerer en næsten uidentificérbar brise som, når den rammer et lokales vindue, vil påvirke indeklimaet i en ekstrem lille grad. Derfor er bare to sammenhængende tidsintervaller med samme værdi for, f.eks., CO₂ ganske usandsynligt uden sensorfejl.

Derfor betragter vi intuitivt sammenhængende tidsintervaller med konstante sensorværdier som fejlmålinger. For at balancere denne antagelse, designes denne subproces til kun at fjerne perioder med konstante værdier hvis der er mere end 5 sammenhængende tidsintervaller med identiske værdier.

Formålet med at fjerne disse perioder er at undgå at vores model(ler) lære mønstre i data der er baseret på fejlrapporteringer: Hvis vores model(ler) lærer konstante værdier er en normalitet, viser vores eksplorative eksperimenter at det har en negativ indvirkning på performance.

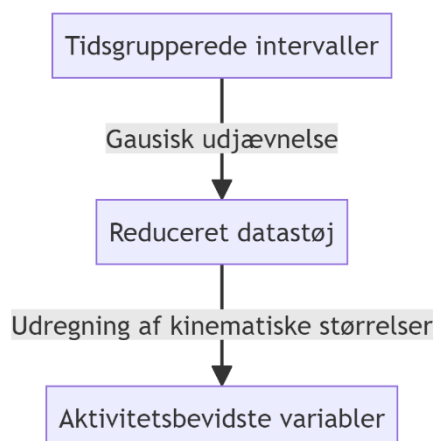
6.3.6 Filtrering af inkonsistente dage

Intuitivt ved vi at vores tidsseriedata indeholder sæsonbestemte og cykliske mønstre, så som typisk anvendelse på tværs af ugedage, skoleferier og forskelle imellem vinter- og sommeraktivitet. Vigtigst for DOMs performance er dog afdækningen af mønstre på dags- og timeniveau. Hvordan ser kurven for, f.eks., CO₂-niveauet ud når et lokale er i brug versus når det ikke er i brug? For at modellere domænet korrekte, må vi sikre os at data ikke instruerer DOM i fejlhæftede eller inkonsistente mønstre. For at sikre dette på dags- og timeniveau, søger vi at fjerne alle dage med en substantiel grad af inkonsistens. Praktisk fjernes alle dage i tilgængelige data hvor sensorværdier for over 25% af de totale tidsintervaller i *dagtimerne* på et givent døgn af tidsintervaller mangler.

6.3.7 Tidsgrupperet databerigelse

Som nævnt ovenfor modellerer vi domænet som tidsserier af sensordata. Intuition bag dette er at (signifikante) skift i sensorværdier imellem sammenhængende tidsintervaller kan hjælpe med at identificere aktivitet. En enkeltstående CO₂-værdi, for eksempel, siger intet om aktivitet, imens en tidsserie af CO₂-værdier viser *udviklingen* i CO₂-niveauet, hvilket er anvendeligt til at identificere aktivitet.

Givet dette design, er det nødvendigt at tage højde for huller i tidsserierne. Hvis en tidsserie, for eksempel, mangler visse intervaller, vil et simpelt rullende gennemsnit springe i tid, og dermed introducere fejlbehæftede forhold imellem sensorværdier i tidsserien. Vi udfører derfor databerigende transformationer på tidsgrupperede intervaller, således at manglende data ikke forårsager at to tidsintervaller der ikke er sammenhængende, influerer hinanden. Under den tidsgrupperede berigelse udføres to subprocesser, vist herunder i figur 26, og gennemgået i efterfølgende punkter. Det overordnede formål med denne subproces er dels at udjævne mindre varianter i tidsserierne, og dels at opstille *aktivitetsbevidste* variable: Variable som repræsenterer udviklingen i en eller flere tidsserier af sensorværdier imellem tidsintervaller.



Figur 26: Proces for tidsgrupperet databerigelse

- **Gaussisk udjævnelse**

Den matematiske udredning af gaussisk udjævnelse (Gaussian smoothing) ligger langt udover omfanget af denne rapport. Gaussisk udjævnelse er, kort forklaret, en form for lavpasfiltrering, der anvendes for at reducere støj og detaljer i navnlig tidsseriedata. Metoden anvender en gaussisk funktion til at skabe en udjævnet version af den oprindelige data ved at give en vægtet gennemsnit af hvert datapunkt og dets naboer, hvilket leder til at små fluktuationer elimineres imens større svingninger forbliver intakte.

I vores transformation anvendes denne metode for at fjerne mindre udsving i sensordata der kunne skyldes at fra at en dør åbnes, over rengøring og gennemgang, til at et varmeapparat skrues én grad op. Omvendt ønsker vi at beholde de større svingninger, som, alt andet lige, bør indikere faktisk aktivitet efter vores definition, præsenteret i baggrunden for DOM.

- **Udregning af kinematiske størrelser**

Begrebet kinematik refererer til den gren af mekanik, der beskæftiger sig med geometrien af objekters bevægelse, så som forskydning, hastighed og acceleration, uden at tage højde for de kræfter, der frembringer sådan en bevægelse. I vores kontekst er dette netop relevant intuition bag tidsseriedata er at (signifikante) skift i sensorværdier imellem sammenhængende tidsintervaller kan hjælpe med at identificere aktivitet, som nævnt ovenfor.

Konkret søger vi at udlede et sæt af kinematiske størrelser (Aktivitetsbevidste variabler) for hver relevant tidsserie af sensordata, givet nedenfor i figur 27. Disse variabler anvendes som input til modellerne for at instruere dem i udviklingen imellem tidsintervaller for relevante tidsserier af sensordata.

Kinematisk størrelse	Definition
Hastighed	Måler ændringsraten af f.eks. CO2-niveauer over tid, hvilket indikerer, hvor hurtigt koncentrationen stiger eller falder i et lokale
Acceleration	Angiver ændringen i hastigheden af f.eks. CO2-niveauændringer pr. tidsinterval, hvilket hjælper med at identificere hvor hurtigt ændringerne i luftkvaliteten forekommer.
Ryk	Måler ændringen i accelerationen af f.eks. CO2-niveauer, hvilket kan indikere pludselige skift eller begivenheder i et lokale, som påvirker luftkvaliteten.
Logaritme	Anvendes til at transformere f.eks. CO2-tidsseriedata, så de store "værdiområder" komprimeres til en mindre skala, hvilket gør det lettere at observere og sammenligne trends og mønstre over tid.

Figur 27: Definitioner af kinematiske størrelser

6.3.8 Tidsrelaterede variabler

Endelig udledes specifikke tidsbaserede variabler for hvert tidsinterval. Denne tilføjelse er resultatet af iterative eksperimenter med tilføjelsen af forskellige variabler for at forbedre den gennemsnitlige performance for de lokalespecifikke modeller. Konkret introducerer vi følgende variabler, præsenteret nedenfor i figur 28.

Variabel	Definition
Nattetime	Hvorvidt tidsintervallet falder i nattetimerne, defineret som værende imellem kl. 00:00 og 06:00.
Ugedagstal	Tidsintervallets tilhørsforhold til den numeriske repræsentation af ugedagene, således at mandag er 0, tirsdag er 1 osv.

Figur 28: Tidsbaserede variabler

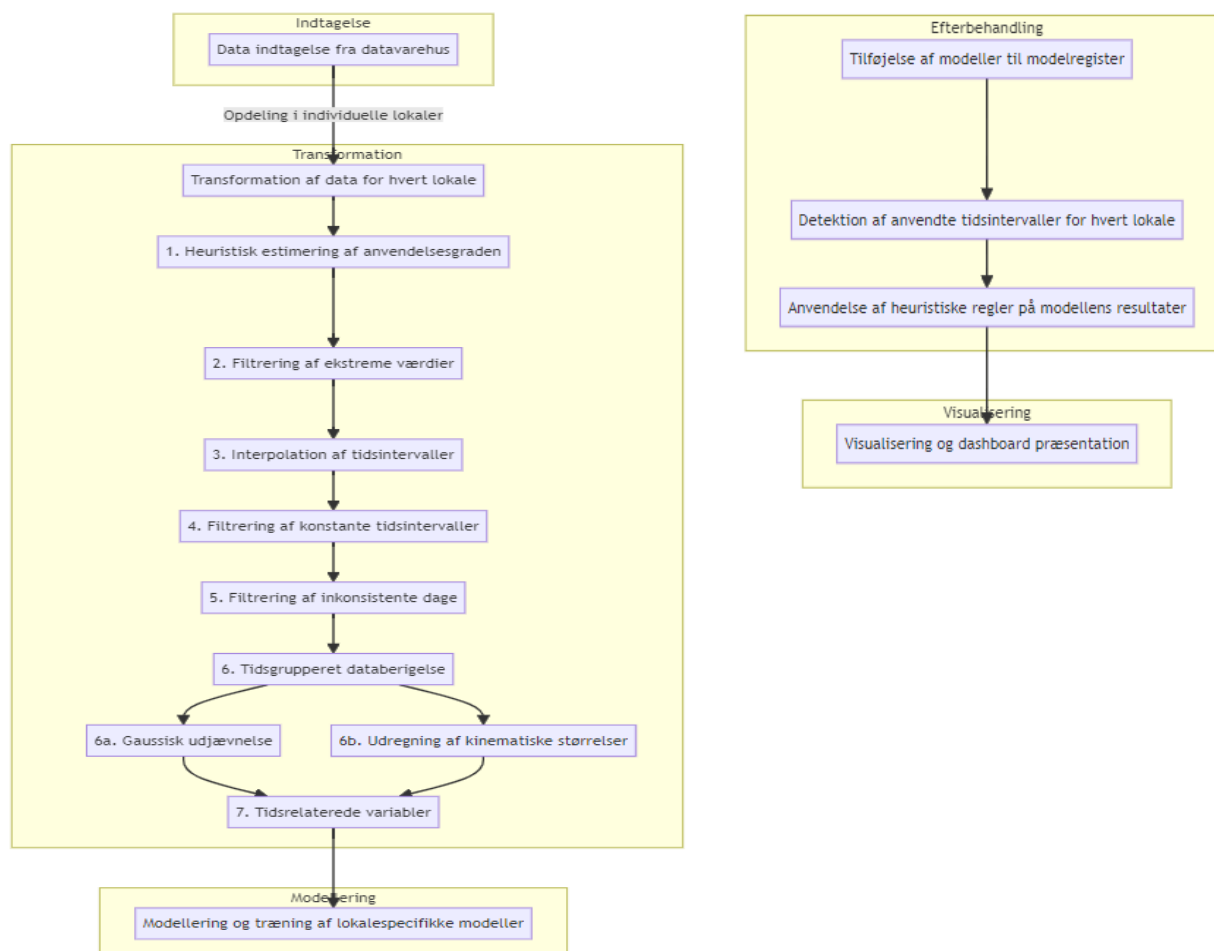
6.3.9 Heuristiske regler

For at introducere yderligere domæneforståelse i løsningen, tilføjes et lag af heuristiske regler der bruges til at augmentere resultaterne fra de lokalespecifikke modeller. Selvom modellerne er i stand til at detektere anvendelse på egen hånd, kan vi forbedre performance ved at kombinere denne statistiske metode med en regelbaseret tilgang. Konkret opstilles følgende regler

1. **Natfiltrering:** Hvis tidsintervallet er mellem kl. 00:00 og kl. 06:00, og anvendelsesscoren er mindre end eller lig med 0,7, vurderes lokalet til ikke at være i brug i det givne tidsinterval. Dette er nyttigt for at minimere falske positive (dvs. situationer hvor en model estimerer at et rum er i brug og det ikke er tilfældet i virkeligheden) i timer, hvor rummet sandsynligvis ikke er i brug. Formål med denne regel er at introducere vores domæneforståelse for aktivitet: Vi ved på forhånd at chancen for at et lokale er i brug i nattetimerne er lav. Derfor tilføjer vi denne regel for at korrigere eventuelle detektioner som modellerne er usikre på.
2. **Enkeltstående tilfælde:** Hvis et givent lokale vurderes som værende i brug i et givent tidsinterval, mens det kronologisk er omgivet af tidsintervaller, hvor lokalet ikke vurderes til at være i brug, ser vi bort fra den givne models vurdering. Intuitionen er baseret på DOMs oprindelige formål: Vi ønsker ikke at identificere kortvarig brug, men kun reelle aktiviteter som undervisning, møder eller fritidsaktiviteter. Denne regel hjælper netop med at fjerne isolerede hændelser, der kan skyldes støj eller kortvarig aktivitet snarere end faktisk rumbrug.
3. **Lave CO2-niveauer:** Hvis CO2-niveauet er mindre end eller lig med 325 PPM, vurderes lokalet altid til ikke at være i brug. Denne regel antager at lave CO2-niveauer indikerer, at et rum ikke er besat, hvilket forbedrer nøjagtigheden af modellernes vurderinger i særlige grænsetilfælde (*edge cases*). Til sammenligning er det global gennemsnit (2023) for CO2 i atmosfæren 421 PPM, hvilket vil sige at vores heuristiske regel sættes lavere end det global gennemsnit.

6.4 Samlet flow

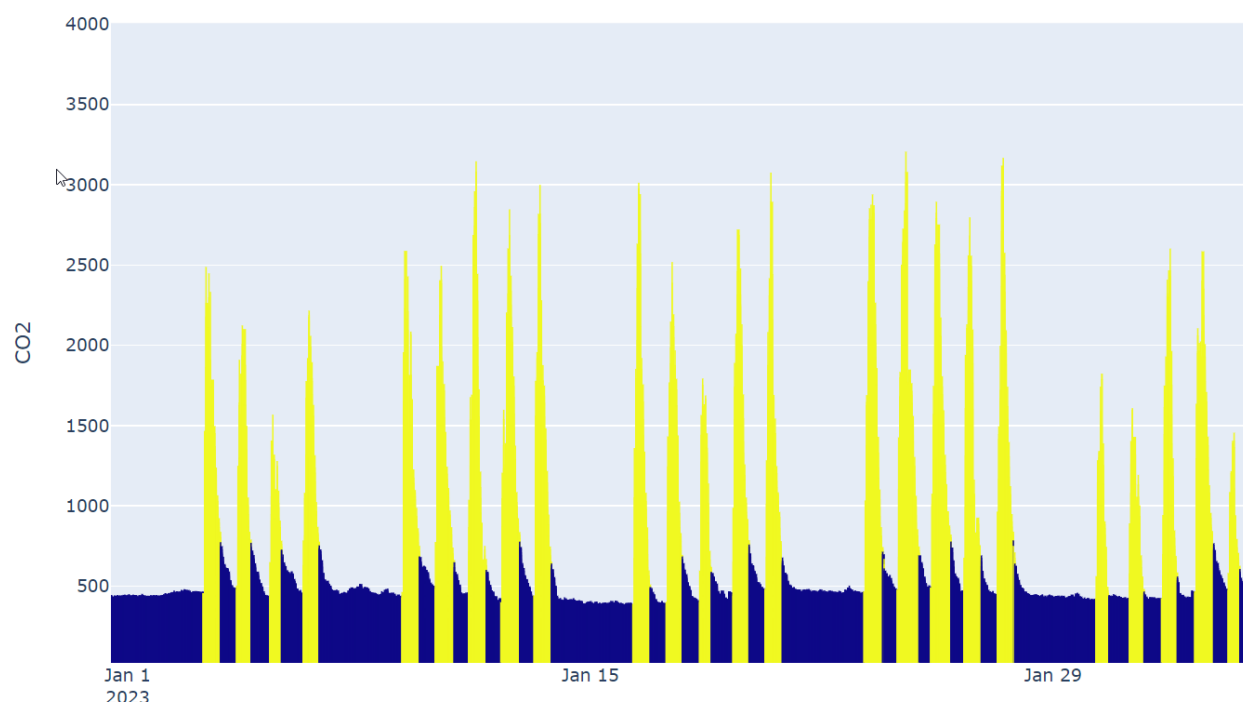
Som beskrevet indledningsvis består løsningens *flow* af en række processer. Disse kan opdeles . Det er vigtigt at understrege at disse processer udføres for hvert lokale i tilgængelige data. Rationalet for dette er gennemgået ovenfor. Som vist nedenfor i figur 30, hentes ny data fra Enformantens datavarehus via Data Factory og opdeles i individuelle lokaler. Derefter transformeres data for hvert lokale, via de syv transformationsprocesser gennemgået ovenfor. Herefter sendes data til modelleringsprocessen hvor lokalespecifikke modeller trænes på den transformerede data. De trænede modeller tilføjes herefter til løsningens *modelregister* under et unikt ID. Samtidigt detekteres alle anvendte tidsintervaller for hvert lokale ved hjælp af lokalets tilhørende model. Derefter anvendes de heuristiske regler på modellens resultater. Det samlede resultat visualiseres (og vises senere via løsningens indbyggede dashboard). Herefter står løsningen klar til at detektere anvendelse på ny data.



Figur 30: Det samlede flow for DOM

6.5 Evaluering og dashboard

Den primære ulempe ved usuperviseret maskinlæring, som nævnt ovenfor, er at kvantitativ evaluering af performance er kompleks. Uden annoteret data, dvs. viden om hvorvidt et lokale rent faktisk var i brug eller ej på et givent tidspunkt, kan vi ikke opstille traditionelle, numeriske metrikker for vores løsning. Givet det relativt korte udviklingsforløb har det ikke været muligt at dedikere projekttid til at søge en kvantificérbar evalueringstrategi. I stedet for har evaluering af modellernes performance været visuelt: Ved at opstille en eller flere tidsserier af sensordata og farvekode de tidsintervaller som modellen for det enkelte lokale har identificeret som værende i brug, har vi en basis for at tilsætte egen domæneforståelse til modellens resultater, og på den måde evaluere performance. Et eksempel på dette vises nedenfor i figur 31.



Figur 31: Visuel inspektion af model performance

Langs den vertikale akse måler vi det egentligt CO2 niveau i PPM. Langs den horisontale akse placerer vi tidsintervaller i kronologisk rækkefølge for et givent rum. Hvis modellen har detekteret at et givent tidsinterval er i brug, farves den tilsvarende søjle gul. Omvendt er søjlen blå hvis lokalet estimeres til ikke at have været i brug. Ved hjælp af denne præsentation af data, kan vi altså analysere udvikling i CO2 og, ved hjælp af vores domæneforståelse, vurdere om det givent tidsinterval ser ud som om det har været i brug. Herefter sammenligner vi med modellens estimer. På den måde kan vi visuelt vurdere hver models performance. For at kunne anvende denne evalueringsmetode i produktion, skal vi have en måde at eksponere disse visualiseringer på. Derfor indeholder Driftsoptimeringsmodellen en indgangsflade (et endpoint) til at tilgå et simpelt dashboard hvorigennem en given bruger vil kunne vurdere performance for hver enkel lokalespecifikke model. Dashboard tilgås på rodnoden af løsningen (<https://app-govtech.azurewebsites.net/>) og er yderligere beskrevet i den tekniske dokumentation. Figur 32 nedenfor viser et skærbillede af dashboardet, med fokus på lokale 02.S.14 på Strandskolen i Aarhus Kommune.



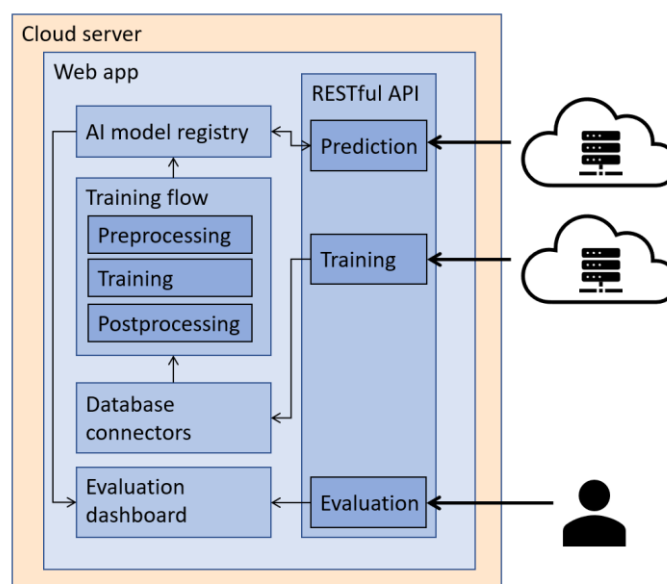
Figur 32: Skærbillede af dashboard-komponenten

6.6 Arkitektur

Processerne beskrevet ovenfor udgør kernen i Driftsoptimeringsmodellen. For at bringe løsningen i produktion, tilføjes et lag af logik der gør det muligt at eksponere DOM som en webapplikation.

Driftsoptimeringsmodellen er designet som en modulariseret mikroservice: En selvstændig applikation med et enkelt formål og en række af moduler der kan modificeres uden indflydelse på resten af arkitekturen. Løsningen er ligeledes designet til at være agnostisk overfor hvilke komponenter den integrerer op imod, i overensstemmelse med AI Signaturprojektets formål. Med ganske få ændringer, vil DOM kunne tilpasses til enhver omkransende arkitektur.

For at forstå den fulde løsningsarkitektur er det nødvendigt at præsentere de involverede komponenter. En visualisering af arkitekturen vises nedenfor i figur 33.



Figur 33: Arkitektur for Driftsoptimeringsmodellen

6.6.1 Modelregister

Fordi anvendelsesgraden ønskes detekteret for hvert lokale, på tværs af skoler, trænes en unik model per lokale, som uddybet ovenfor. Med udgangspunkt i den tilgængelige data resulterer dette i 150+ AI-modeller der til sammen trænes på over 9 mio. dataværdier. Denne mængde modeller gør det nødvendigt at skabe et modelregister for systematisk at kunne tilgå og anvende hver model i den rigtige kontekst. Her anvendes en ikke-relationel in-memory datastruktur i hvilken hvert lokale, for hver skole, indekseres under et unikt ID, udledt af skolen og lokalets navn. Når løsningen instrueres til at vurdere ny data, kaldes den relevante model fra modelregisteret og detekterer anvendte tidsintervaller for det tilsvarende lokale.

6.6.2 Integrationer og autentifikation

I produktion kan slutbrugere og/eller automatiserede pipelines interagere med DOM igennem tre *endpoints*. Systemet er designet til at modtage instrukser om at påbegynde træning (*/train*) eller detektion (*/predict*), og derefter hente den relevante data fra cloudbaseret dataopbevaring (Fra Snowflake via Azure Data Factory i vores projektarkitektur). Herudover indeholder løsningen en dashboard-komponent som gennemgås i et særskilt afsnit nedenfor.

For at interagere med løsningen kræves en OAuth2-autentifikation der udløser en tidsbegrænset *token* med rettigheder til at tilgå systemets funktionaliteter. For at kunne anmode om denne autentifikation skal brugere og/eller automatiserede pipelines optræde på en prædefineret liste af godkendte brugere. På den måde tilføjes et sikkerhedslag til løsningen.

En fuld oversigt over integrationer samt specifikationer for autentifikation er tilgængelig i den tekniske dokumentation.

6.7 Konklusion

Samlet set er Driftsoptimeringsmodellen en umoden men lovende løsning der i grove træk løser det opstillede formål. Siden datagrundlaget ikke indeholder viden om hvorvidt et givent lokale rent faktisk var i brug eller ej, har evalueringsprocessen for de rumspecifikke modellers performance været udelukkende visuel. Dette betyder at vi endnu mangler robuste, kvantitative metrikker for løsningens performance, selvom man ved visuel inspektion tydeligt ser at de underliggende AI-modeller – i langt de fleste tilfælde – er i stand til at differentiere imellem lokaler der er i brug og lokaler der ikke er i brug. Selve arkitekturen giver rig mulighed for at videreudvikle løsningen, blandt andet ved at højne dataintegriteten eller at introducere annoteret data. I sin nuværende form er Driftsoptimeringsmodellen klar til at gå i produktion, mens der bør understreges at der stadig er vigtige aspekter af løsninger der bør adresseres på kort eller mellem sigt. Bemærk at forslag til videreudvikling optræder i det sidste afsnit i rapporten.

7. Energibevidst AI Booking

Den følgende sektion forklarer Energibevidst Bookingsystem (EAB i følgende; Endnu unavngivet): Et AI-innovationsprojekt til at augmentere lokalebookingprocessen med information om energioptimale valg, udviklet under AI signaturprojektet *Optimeret Bygningsanvendelse*. Bemærk at brugsdokumentationen for løsningen er at finde under den supplerende tekniske dokumentation.¹

7.1 Baggrund

Formålet med OPI-projektet er overordnet at opnå besparelser på energi- og klimaforbrug ved at konsolidere aktiviteter, implementere intelligent lokal allokering og øge udnyttelsesgraden af kommunens bygningsmasse ved brug af bygningsdata, IoT-løsninger (Internet of Things) og kunstig intelligens (AI). Baggrunden for denne oprindeligt indgående OPI-aftale om et AI Signaturprojekt med NTT DATA, var den fælles interesse i at anvende AI (og relaterede moderne teknologier) til at adressere ovenstående i en innovationsorienteret, eksplorativ arena. Projektet "*Energibevidst AI Booking*" (EAB) er baseret på denne tilgang, og er som følge et innovationsprojekt der, ved projektets afslutning, fortsat kræver yderligere udvikling og supporterende arkitektur for at kunne operationaliseres hos de deltagende kommuner.

7.2 Overblik over EAB

EAB-løsningen, eller Energibevidst AI Booking, søger at strømline brugen af energi i kommunale bygninger ved at hjælpe fritidsbrugere med at tage energibevidste valg i en bookingsituation. Formålet med systemet er at tilbyde en praktisk, data-drevet metode til at optimere booking af lokaler, med direkte indvirkning på energiforbruget.

Løsningens kernefunktionalitet er dens evne til at foreslå den mest energioptimale næste booking i realtid, ved at tage højde for eksisterende bookinger, energidiffusion imellem lokaler, energiforbrug og anvendelsesgrad i forskellige dele af en given bygning. Med udgangspunkt i eksisterende bookinger og opvarmningsmønstre i bygningens fløje og etager, søger systemet at angive hvilket lokaler der vil være det optimale at booke næste gang.

Hvis en del af bygningen, for eksempel, allerede er opvarmet (eller nedkølet) grundet tidligere eller løbende bookinger, søger systemet at favorisere bookinger i denne del for at undgå unødvendig opvarmning af andre områder. Dette er særligt relevant i store bygninger med flere fløje og etager, hvor opvarmningsbehovet kan variere betydeligt.

¹ Tilgængelig på github.com/Enformanten/ai-booking-system

Udgangspunktet for EAB er at anvende bygningsplaner, bookinghistorisk og målt energiforbrug til at informere hvor og hvornår den næste booking burde placeres. Rumkapacitet og rumfaciliteter tages også i betragtning, idet nogle aktiviteter kræver specifikke faciliteter, så som køkkeninventar, instrumenter, whiteboards, fjernsyn, projektorer osv. Vision er, at de tiltænkte brugere af systemet, fra bygningsforvaltere til fritidsbrugere, vil kunne træffe informerede, energi- og klimabevidste beslutninger, når de booker ved hjælp af realtidsinformationen fra AI-systemet. Ud fra dette perspektiv udforsker vi i dette projekt om de deltagende kommuner kan opnå besparelser på energiforbrug og samtidig sikre en mere effektiv anvendelse af deres bygningsmasse.

7.3 Proces

EAB blev konceptualiseret allerede inden OPI-aftalens indgåelse som et bærende element i samarbejdet, under konsensus af at dette initiativ var innovativt i natur. Formålet var altså ikke frembringe et produktionsklart system, men at udforske *hvad der var muligt* indenfor den ramme der fremsættes ovenfor. Derfor har projektet været domineret af dataanalyse og eksperimentel modellering i en iterativ cyklus med målet om at finde et gunstigt design baseret på tilgængelige data eller data der rimeligvis burde kunne tilføres i fremtiden.

7.4 Data og associerede begrænsninger

Idet signaturprojektet er et samarbejde imellem de deltagende kommuner om at finde tværgående løsninger, følger det naturligt at vi ønsker at anvende datakilder som er konsistente på tværs af kommunernes datalandskab således at vi kan designe en løsning der kan bruges af alle. Denne præmis begrænser intuitivt de datakilder vi kan designe en løsning på baggrund af, men forhindrer ikke vores eksplorative arbejde. EAB blev konceptualiseret endnu før projektgruppen besad en fyldestgørende forståelse for hvilket datagrundlag der kunne forventes. Ligesom for Driftoptimeringsmodellen har vi tre kommuner, som hver indeholder én skole med et variabelt antal lokaler. For en delmængde af hver skoles lokaler har vi adgang til en multivariat tidsserie af sensordata, som beskrevet ovenfor og vist i figur 34 nedenfor. Indenfor dette domæne har vi ligeledes adgang til en mængde metadata for hver skole og lokale, hvorunder særligt historiske bookinger, energiforbrug på bygningsniveau og bygningsplaner er relevante for dette projekt.

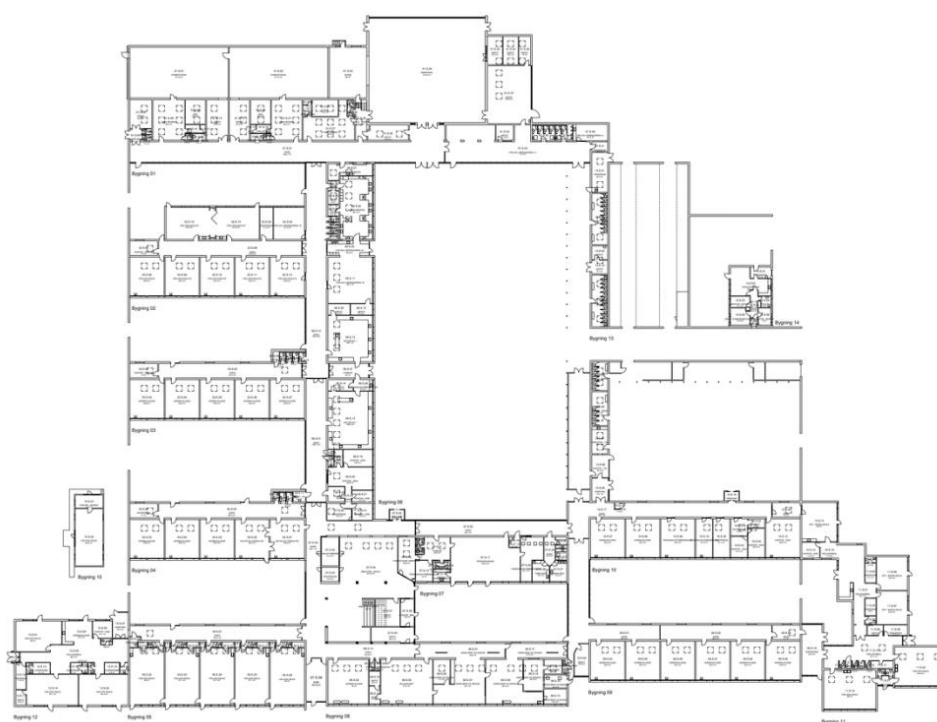
Rum ID	Dato	Tidspunkt	Temperatur	CO2	Kommune	Skole
12.1.002	07-09-2022	10:45:00	21.8	997.33	Syddjurs	Thorsager Skole
12.1.002	07-09-2022	11:00:00	21.73	872	Syddjurs	Thorsager Skole
12.1.002	07-09-2022	11:15:00	21.83	905.33	Syddjurs	Thorsager Skole
12.1.002	07-09-2022	11:30:00	22	937	Syddjurs	Thorsager Skole
12.1.002	07-09-2022	11:45:00	21.86	602.33	Syddjurs	Thorsager Skole
12.1.002	07-09-2022	12:00:00	21.9	663	Syddjurs	Thorsager Skole

Figur 34: Uddrag af rådata, lokalespecifikke tidsserier

Fordi den grundlæggende målsætning er at kunne vurdere hvilket lokale der fortrinsvis er mest energiefficient at booke baseret på den øjeblikkelige kontekst, søges vi at frembringe et datasæt med flere attributter for hvert lokale der tilsammen danner et helhedsbillede. Disse behandles separat nedenfor.

7.4.1 Lokalets placering

Modsat Driftsoptimeringsmodellen, er vi i dette projekt nødsaget til at forstå omgivelserne for et givent lokale, for at forstå et lokales energioptimalitet. For eksempel følger det intuitivt at fysiske love om termisk inertie at energidiffusion vil spille en rolle for vores eksperimenter: Hvis lokale A og lokale B deler en væg, vil opvarmning af lokalet A påvirke temperaturen i lokale B og vice versa i en grad bestemt af energiniveau og isolering. Hvis lokale A er booket og derfor opvarmet vil det, alt andet lige, været billigere at varme lokale B op, hvilket er en nyttig indsigt for vores optimeringsalgoritme i EAB. Forholdet imellem lokaler i større bygningsmasser er naturligvis langt mere komplekse. En reel bygning kan f.eks. have flere etager, tværgående gange og variationer i soleksponering, varme- og energikilder og områdespecifikke ventilationssystemer. Figur 34 (Nedenfor) viser en bygningsplan, som eksempel på denne kompleksitet.



Figur 35: Bygningsplan (Strandskolen, Aarhus Kommune)

7.4.2 Lokalets brugshistorik

Projektets formål nødvendiggør viden om hvordan et givent lokale bliver brugt, både for at kunne simulere en bookingsituation og for at kunne sammenholde energiforbrug med aktivitet. Til formålet har vi adgang til

historiske bookinger for fritidsbrugere (hentet fra flere forskellige systemer), men ikke for brug i skoletimerne. Figur 34 nedenfor viser et uddrag af bookinghistorik udenfor skoletiden for lokaler på Strandskolen i Aarhus Kommune.

Rum ID	Dato	Tidspunkt (hel time)	bookingstatus	Kommune	Skole
06.S.02	11-08-2022	17:00:00	1	Aarhus	Strandskolen
01.S.37	11-08-2022	17:00:00	1	Aarhus	Strandskolen
01.S.01	11-08-2022	17:00:00	0	Aarhus	Strandskolen

Figur 36: Uddrag af rådata, bookinger

7.4.3 Lokalets faciliteter

Zoomer vi ud fra vores betragtninger om energiforbrug og bookinghistorik, og betragter EAB i en praktisk kontekst, må vi også tage højde for et givent lokalets *faciliteter*, så som borde, stole, tavler, whiteboards, projektorer osv. Det kan være nødvendigt for en "booker" at et lokale har en eller flere specifikke faciliteter. Det må antages, eksempelvis, at klaverundervisning kun kan finde sted i et lokale med et klaver. En madlavningsklub skal ligeledes bruge et køkken. Mere generelt set kan det være nødvendigt at have adgang til en projektor eller et fjernsyn. Fritidsbrugere vil altså booke et lokale til et specifikt formål, hvilket betyder at kun specifikke lokaler med de rigtige faciliteter er relevante. I projektet har det ikke været muligt at inddrive detaljerede oplysninger om faciliteter i lokalerne. Dette skyldes primært at informationerne ikke var umiddelbart tilgængelige for de deltagende kommuner og at udfordringerne omkring at tilegne sig mere essentielle data, så som bookinghistorik og energiforbrug, blev forfordelt.

7.4.4 Lokalets energiforbrug

Den grundlæggende præmis for at foretage energibevidste valg i lokalebooking er en viden om hvert lokales energiforbrug. Til projektet har vi adgang til informationer om energiforbrug på *byggningsniveau* og på *målerniveau*, hvilket kræver en segmentering – en regression – af energiforbruget ned på aktiviteten for et givent lokale. Figur 35, nedenfor, viser et uddrag af den tilgængelige data.

Måler ID	Dato	Type	Værdi	Enhed	Tidspunkt (hel time)	Dato	Kommune	Skole
75	25-01-2022	Fjernvarme	0,16	MWh	12:00:00	2022-01-25	Favrskov	Rønbækskolen
75	25-01-2022	Fjernvarme	0,17	MWh	13:00:00	2022-01-25	Favrskov	Rønbækskolen
75	25-01-2022	Fjernvarme	0,13	MWh	14:00:00	2022-01-25	Favrskov	Rønbækskolen
75	25-01-2022	Fjernvarme	0,10	MWh	15:00:00	2022-01-25	Favrskov	Rønbækskolen
75	25-01-2022	Fjernvarme	0,08	MWh	16:00:00	2022-01-25	Favrskov	Rønbækskolen

Figur 35: Uddrag af rådata, energivariabler

7.5 Tilgang og designvalg

Fra det opstillede domæne og ovenstående kan vi identificere fire afgørende faktorer som danner grundlaget for vores tilgang til eksperimentering under rammen af EAB, opstillet nedenfor, med de primære afledte designvalg for hver faktor.

- ***Målet er at kunne anbefale det mest energioptimale, relevante lokale at booke i øjeblikket***

Den overordnede præmis for EABs design er at kunne *anbefale* hvilke bookingmuligheder (Det vil sige ville lokaler på hvilket tidspunkt) der er mest energioptimale i øjeblikket. Grundlæggende kan vi altså betragte EAB som et kontekstbevidst anbefalingssystem: En løsning der, ved at se på konteksten (Så som hvilke andre lokaler der er booket), kan vurdere hvilken bookingmuligheder der er at foretrække.

Vi kan også udlede at energioptimalitet ikke er den eneste faktor der er vigtig i en bookingsituation: En "booker" vil have visse krav til en bookingmulighed så som lokalets kapacitet og faciliteter, som nævnt indledningsvis. Der er altså flere typer *omkostninger* (energiorienterede eller brugsorienterede) at tage højde for. Det betyder praktisk at en opstillet omkostningsfunktion må have et antal termer lig med antallet af relevante omkostninger.

- ***Tilgængelige data er en ufuldkommen repræsentation af virkeligheden***

Selvom vi besidder f.eks. plantegninger for de relevante skoler, er der stadig mange uopklarede spørgsmål om skolernes og lokalernes attributter. Spørgsmål som "*hvordan er det enkelte lokale isoleret*" eller "*hvilke materialer er skolen bygget med*" er blot starten på det omfattende eksplorative arbejde der ville skulle udføres for at kunne vurdere f.eks. varmediffusion korrekt. Ligeledes er vores forståelse for ventilation og det generelle energiforbrug for hver bygning minimal. Den fundamentale problematik i at modellere domænet er således ikke designet af en tilstrækkeligt repræsentativ algoritme, men et spørgsmål om tilgang til information, hvilket har været et gennemgående problem for signaturprojektet. Der har fra konceptualiseringen af EAB været en forståelse for at visse simplificerende antagelser måtte laves for at kunne arbejde med dette domæne. I projektets forløb er antallet af simplificerende antagelser vokset støt i takt med at kompleksiteten af domænet er blevet afdækket.

- ***Løsningsdesignet må bygges under antagelser om data***

Kort inde i projektføreløbet stod det klart at datagrundlaget ikke ville kunne danne en fuldkommen repræsentation af domænet (som gennemgået ovenfor). Dertil kommer, at visse elementer er kritisk natur for projektet, er utilgængelige indenfor tidsrammen af AI Signaturprojektet. F.eks. vil brugsorienterede begrænsninger (Så som et lokales faciliteter eller kapacitet) ikke være mulige at få adgang til. Fordi disse

informationer eksisterer, men blot er uden for rækkevidde i projektet, kan vi retfærdiggøre at introducere *syntetisk data*. Vi vil anvende syntetisk data til at dække de huller i datagrundlaget som der antagelsesvis må være muligt at lukke i en senere fase.

- ***Tilgængelige data varierer på tværs af kommuner og skoler***

Som nævnt ovenfor, er det en præmis for hele AI signaturprojektet at vi tilstræber at konstruere fælles løsninger på tværs af datalandskaber i de involverede kommuner. Fordi mængden, integriteten og typerne af tilgængelige data fra hver kommune varierer, søger vi i EAB at gøre brug af typer af sensordata der er konsistente på tværs af kommunerne. Dette indskrænker naturligvis vores råderum for inddragelse af interessante datakilder så som ventilationsinformationer, som desværre kun er tilgængelige for en delmængde af skolerne.

7.5.1 Samlede designvalg

Samlet set søger vi altså at opstille et *anbefalingssystem* baseret på en *aggregeret omkostningsfunktion* der kan *prissætte* den *relative* (regressive) energi- og brugsomkostning for hvert lokale og rangere disse i *realtid*. Den anvendte data skal være *tilgængelig for alle involverede kommuner/skoler* eller *syntetisk* i det omfang det kræves for at opstille en minimal *datarepræsentation* af domænet.

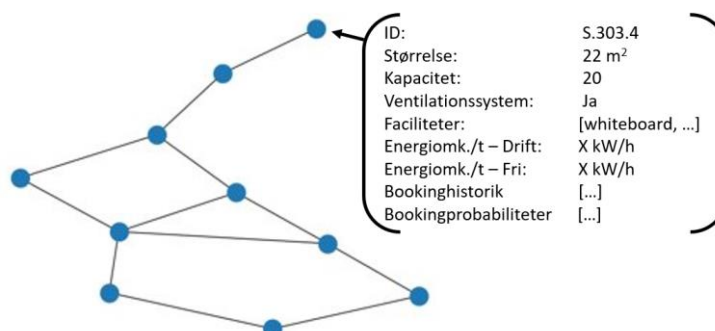
7.6 Eksperimenter

EAB kan segmenteres ned i tre relaterede eksperimenter: Bookingsystemet, Omkostningsberegningen og Energiestimering. Disse gennemgås individuelt nedenfor. Det bør holdes i mente, at det ovenfor nævnte retningsskift i AI Signaturprojektet resulterede i et abrupt ophold i design- og udviklings af disse eksperimenter. Siden er kildekoden bag EAB blevet renskrevet med det formål at berede projektet på videreudvikling i et efterfølgende forløb.

7.6.1 Omkostningsberegning

En bygning kan betragtes som et netværk, hvor hvert lokale (og andre rum, f.eks. gange) repræsenteres som en node og deres inter partes sammenhæng repræsenteres som forbindelser imellem noder. En netværksrepræsentation giver os et matematisk fundament (Se den tekniske dokumentation) for vores eksperimenter.

At repræsentere lokaler som noder muliggør en objektbetragtning, hvor hvert lokale kan tilskrives et sæt af attributter så som lokalets størrelse, kapacitet, faciliteter, bookinghistorik og, endnu vigtigere, dets relative placering i forhold til andre lokaler og rum – Altså nodens forbindelser til andre noder. Et syntetisk eksempel gives nedenfor i figur 36.



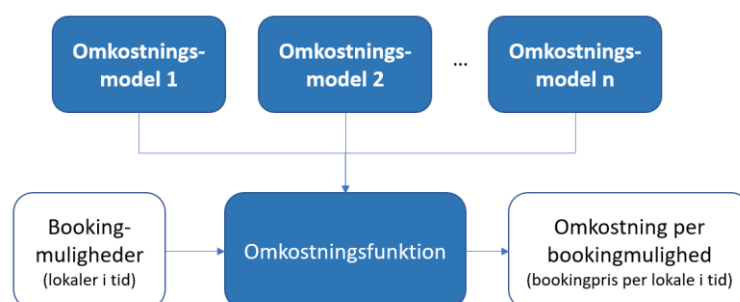
Figur 36: Grafrepræsentation: Syntetisk bygning med 10 rum

Grafrepræsentationen er særlig vigtig fordi den søgte omkostningsberegning ikke kan repræsenteres lineært, men derimod vil være afhængig af afledte præsent forholds i resten af bygningen, udendørstemperaturen, bookingstatus i omkringliggende lokaler osv. Med denne indsigt i mente, opstiller vi et sæt af initiale omkostningstyper som vil udgøre subelementerne i vores omkostningsfunktion. Disse er vist nedenfor i figur 37.

Omkostningstype	Beskrivelse	Område
Energi - Naboskab	Siden lokaler ikke opvarmes enkeltvis men i sektioner, er bookingstatus i tilstødende lokaler aksiomatisk vigtigt for omkostningsberegning. Det antages her, at det er billigere at booke lokaler i allerede opvarmede sektioner.	Energi (realomkostning)
Energi - Regression	Ved at forstå den regressive sammenhæng imellem det enkelte lokales brug og det samlede energiforbrug for bygningen, søger vi at tillægge hvert lokale en kontekstafhængig energikoefficient.	Energi (realomkostning)
Kapacitet	Et lokales relevans i en given bookingsituation er dikteret af dets kapacitet relativt til antallet af individer lokalet bookes til. Det antages her at relevansen af lokaler falder som forskellen imellem kapacitet og antal individer vokser.	Relevans for booking
Faciliteter	Et lokales relevans i en given bookingsituation er dikteret af dets faciliteter relativt til hvilke faciliteter der efterspørges. Det antages her at relevansen af et lokale falder som forskellen imellem faciliteter og efterspurgte faciliteter vokser.	Relevans for booking

Figur 37: Inkluderede omkostningstyper

For hver af de ovenstående omkostningstyper opstilles et subeksperiment af varierende kompleksitet. Disse eksperimenter omsættes derefter til selvstændige, isolerede *omkostningsmodeller* der tildeler hver bookingmulighed en kontekstbevidst omkostning på bookingtidspunktet. For hvert lokale summeres disse værdier til en samlet lokalespecifik omkostning. Dette flow er vist nedenfor i figur 38.



Figur 38: Proces for omkostningsvurdering per bookingmulighed

Omkostningstyperne "Kapacitet" og "Faciliteter" behandles samlet under *Bookingsystemet* mens "Energi - Naboskab" og "Energi – Regression" behandles under *Energiestimering* i følgende sektion. Det bør igen noteres, at projektets abrupte afslutning forhindrede videre udvikling af nye omkostningstyper og grundig optimering af de ovenstående. Foreslag til et videre forløb behandles afsluttende under Fremtidige perspektiver.

7.6.2 Energiestimering

Som nævnt ovenfor har vi to energirelaterede omkostningsmodeller ved projektets afslutning. Disse behandles særskilt nedenfor.

7.6.3 Energi – Naboskab

VVS (Varme-, Ventilations- og Sanitetsteknik) er et opfattende emne med enorm relevans for designet af en løsning som EAB. Som beskrevet ovenfor er både projektets datagrundlag og vores domæneekspertise indenfor dette emne ikke dybdegående nok til at kunne danne en fyldestgørende repræsentation. Derfor beror vores tidlige omkostningsmodel for opvarmning i stedet på et simpelt aksiomatisk princip om diffusion: den proces, hvorved energi spredes inden for et materiale eller mellem forskellige materialer, der er i kontakt, på grund af molekylernes termiske bevægelse.

Fra vores løbende dialog med domæneeksperter ved vi at lokaler ikke opvarmes individuelt: I stedet tilføres varme på et sektionniveau, således at en fløj eller gruppe af lokaler opvarmes samtidig. Ud fra denne indsigt, og med vores netværksrepræsentation af en given bygning, kan vi lave den simple antagelse at *hvis et lokale er opvarmet, er det fordelagtigt at booke lokalet ved siden af*. På den måde vil vi altså kunne undgå at opvarme nye fløje eller sektioner af en bygning.

Konkret bruger en varmediffusionsmodel til at simulere varmeoverførsel mellem rum og til skolens ydre. Her laver vi en serie af simplificerende antagelser som ikke desto mindre er relevante for videreudvikling.

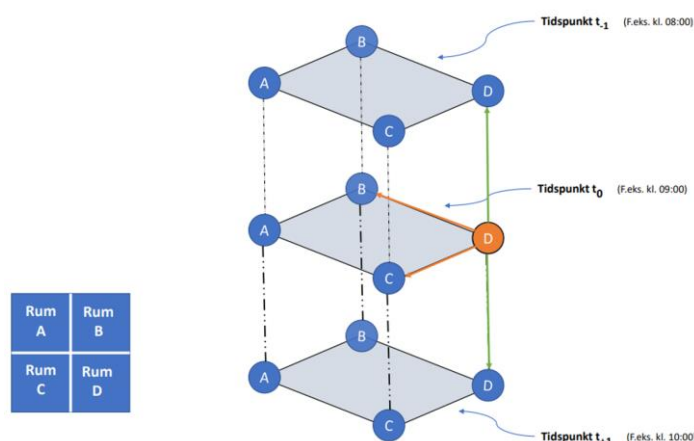
- De antages, at bygningen og dens indre vægge er godt isolerede sammenlignet med luftens termiske ledningsevne, dvs. at temperaturen i et rum er næsten konstant i forhold til positionen inden i det.
- Der vil være strålingseffekter (det vil sige, rum, der vender mod syd, vil være varmere end rum, der vender mod nord, når opvarmningen er slukket), men, givet projektets omfang, er disse ikke taget i betragtning.
- Der vil være konvektionseffekter, dvs. varmetab på grund af luftstrømme, både fra mennesker, der åbner og lukker vinduer og fra bygningens obligatoriske ventilation. Modellering af dette ligger langt udover omfanget af dette projekt, i hvert med den tilgængelige information. Vi har ikke inkluderet en konvektionsterm i vores matematiske repræsentation, da det ville gøre simuleringen signifikant mere kompleks, og vi håber, at det termiske tab på grund af indbygget ventilation kan assimileres til diffusions tab gennem væggene.
- Det antages, at skolen har en form for automatisering indbygget i varmeapparaterne, så de kan tænde, når der er undervisning, og slukke, når rummet ikke bruges. Bemærk, at vi ikke har foretaget en fuld undersøgelse af, hvordan en egentlig styring ville fungere, da vi ikke har overvejet *retrofitting*-effekter i styringen.
- Der er ikke taget hensyn til andre varmekilder, såsom mennesker i rummet, lyspærer til belysning eller computere. Derfor antager vi, at rummet er forholdsvis stort sammenlignet med antallet af mennesker "klemmt ind".

Disse overvejelser resulterer i et lineært inhomogent system af almindelige differentialligninger. Selvom vi ved, at det ville være muligt at omdanne problemet til Laplace-rummet for at få et system af lineære algebraiske ligninger i stedet og derefter transformere tilbage, hvilket gør modulet mindre beregningstungt, har vi valgt ikke at gøre dette for øjeblikket. I stedet vælger vi at bruge en *Runge-Kutta*-metode, som er mere intuitiv og nemmere at udvide. Det samme gælder for integralen til at beregne den kumulative energi: hvor Simpson-metoden anvendes, fordi den er mere fleksibel.

Vi approksimerer at omkostningen ved at opvarme et givet lokale er lig med omkostningen ved at opvarme det lokale, som om det var det eneste lokale, der blev brugt den dag, minus de energibesparelser, der kommer fra andre lokaler, der allerede er booket i nærheden. Vi antager, at disse energibesparelser kommer fra tre kilder:

- Opvarmning af et lokale, der deler en væg i samme tidsrum med et lokale, der allerede er i brug,
- Lokalet er allerede varmt, fordi det er blevet brugt i det foregående tidsrum,
- Hvis dette lokale er booket, vil det allerede være varmt, når det bruges i det næste tidsrum.

Intuitivt kan vi tænke på denne model som hvert lokale har en iboende omkostning og derefter modtager beskeder fra sine nærmeste naboer i netværksrepræsentationen af bygningen om at reducere lokalets omkostning, hvis de allerede er booket. De optimale lokaler at booke vil altså være dem, der modtager flest (eller de vigtigste) beskeder, og dermed har den laveste omkostning. Dette visualiseres nedenfor i figur 39.



Figur 39: Visualisering af omkostningsberegning baseret på naboskab i syntetisk bygning

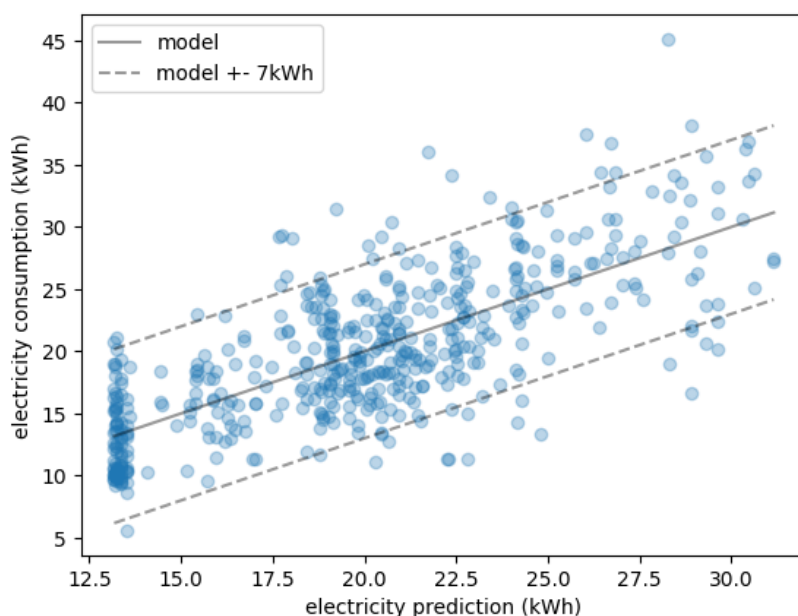
I figur 39 visualiserer vi en bookingsituation hvor kun lokale D er booket på tidspunkt 0, dvs. kl. 09:00. Fra ovenstående ved vi at det herfra er fortrinsvis fordelagtigt at booke lokale D igen på tidspunkt -1 (kl. 08:00) eller på tidspunkt 1 (kl. 10:00). Disse to bookingmuligheder vil være lige energioptimale i dette eksempel. At booke lokale B eller C på tidspunkt 0 (kl. 09:00) er de næstbedste muligheder ud fra den opstillede omkostningsmodel.

7.6.4 Energi – Regression

Den næste omkostningsmodel fokuserer på modellering af den samlede energiomkostning ud fra elektricitetsforbrug. I modsætning til den ovenstående model, der primært er baseret på teoretiske principper for varmediffusion, er denne model datadrevet. Den udnytter historiske data om booking og energiforbrug på bygningsniveau for at etablere en regression, der kan approksimere et givnet lokales indflydelse på det samlede energiforbrug på et specifikt tidspunkt.

Ved at analysere denne bookinghistorik og sammenkøre den med energiforbrugsdata, søger vi altså at opstille en omkostningsmodel der kan forudsige, hvordan forskellige bookinger påvirker bygningens samlede energiforbrug. Målet er at identificere mønstre og tendenser, som kan hjælpe os med at forstå, hvilke lokaler og bookingmuligheder der historisk er mest energieffektive. Praktisk søger vi at udtrække energikoefficienter for hvert bookingmulighed, dvs. hvert lokale på hvert tidspunkt, for at kunne anvende disse i vores omkostningsmodel som *vægtninger* for hver bookingmulighed. Udvalgte visualiseringer fra den bagvedliggende eksplorative dataanalyse kan findes i appendiks A.

Til formålet opstilles en serie af regressionsmodeller med det formål at etablere en sammenhæng mellem rumbrug og energibehov. Figur 40 nedenfor viser performance for en parametriseret *ElasticNet* linear regression.



Figur 40: Modelperformance, lineær regression (Strandskolen, Aarhus Kommune, 2022)

I Figur 40 repræsenterer den midterste linje modellens forudsigelser, som indikerer sammenhængen, modellen har fundet mellem de anvendte vægtninger og elektricitetsforbrug. Ved en perfekt performance vil alle datapunkter (Individuelle forudsigelser versus faktisk forbrug) befinde sig på denne linje.

De stiplede linjer repræsenterer en heuristisk fejlmargen omkring modellens forudsigelser—specifikt er disse linjer placeret plus eller minus 7 kWh fra modellens "forudsigelseslinje". Dette indikerer det område, hvor modellens forudsigelser typisk falder indenfor. Vi ser at modellen har en rimelig nøjagtighed – Op mod 95% af forudsigelserne falder indenfor vores konfidensinterval. Spredningen antyder dog variabilitet i modellens performance: Nogle punkter falder uden for 7 kWh-marginen, hvilket indikerer tilfælde hvor modellens forudsigelser var mindre præcise. Numerisk opnår vi en R^2 score på 0.52 (+/- 0.03).

Under de simplificerende antagelser (Se forrige afsnit), er vi derved i stand til at opstille en relativ pålidelig regressionsmodel for lokalespecifikt (bookingmulighedsspecifikt) energiforbrug. Vi kan derfor bruge de afledte modelkoefficienter som basis for en vægtning i en datadrevet omkostningsfunktion i vores samlede omkostningsfunktion.

7.7 Bookingsystemet

EAB tager, som beskrevet ovenfor, afsæt i omkostningsmodeller indenfor både energi (realomkostning) og bookingrelevans. Til sidstnævnte kategori, så vi i figur 37 at det initiale omfang af vores eksperimenter

inkluderer omkostningsmodeller for *kapacitet* (Antallet af individer et lokale er normeret til) og *faciliteter* (De faciliteter et lokale har). Matematisk følger begge omkostningsmodeller en *Lennard-Jones potential*-funktion.

7.7.1 Kapacitet

Praktisk set antager denne omkostningsmodel at man angiver hvor mange individer man booker til i bookingsituationen.

Når vi overvejer kapaciteten af et lokale til en aktivitet, er det vigtigt at finde den rette balance. På den ene side bør vi ikke anbefale et lokale, der er for lille. Hvis deltagerne ikke vil kunne passe ind, er effektiviteten af opvarmning eller belysning irrelevant. På den anden side bør forslag om lokaler, der er meget større end nødvendigt, undgås. Det kan ikke alene spilde energi ved at opvarme et stort rum for kun få personer, men det betyder også, at en større gruppe senere kunne gå glip af at bruge rummet mere effektivt. For at adressere dette, opstiller vi en omkostning for mismatches imellem et lokales kapacitet og det antal individer der bookes til, som supplerer de andre energiomkostninger forbundet med elektricitet og opvarmning.

Vores omkostningsfunktion for kapacitet tager højde for antallet af deltagere. Hvis antallet af deltagere er mindre end lokalets kapacitet, kan lokalet ikke bookes – det er ganske enkelt for småt. Hvis antallet af deltagere præcis matcher lokalets kapacitet, er der ingen ekstra omkostning. Men hvis der er ekstra plads, indfører vi en yderligere omkostning for hvert ubesat sæde.

Denne yderligere omkostning er skaleret i forhold til lokalets kapacitet. For eksempel er overskydende plads i et mindre lokale mere betydningsfuld end det samme antal tomme pladser i et større lokale. Et lokale beregnet til ti personer, der er halvtomt, er mindre ideelt end et lokale til hundrede personer med fem ledige sæder. Den ekstra omkostning justeres derfor tilsvarende, mindre for det større lokale og større for det mindre lokale, hvilket afspejler den relative vigtighed af at fylde hvert sæde. På denne måde kan vi finjustere omkostningsfunktionen for hver skole, hvilket sikrer, at lokalerne anvendes så effektivt som muligt.

7.7.2 Faciliteter

Ligesom i kapacitetsomkostningsmodellen, antager denne omkostningsmodel at man angiver hvilke faciliteter man har behov for i bookingsituationen.

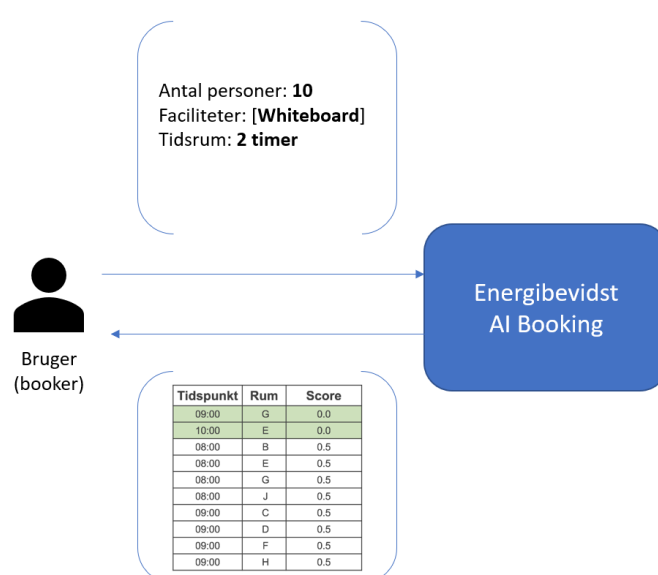
Når man vælger et lokale til en aktivitet, spiller faciliteterne i lokalet – såsom en skærmprojektor, et whiteboard osv. – en afgørende rolle. På samme måde som med kapacitetsomkostningsmodellen, hvor for små lokaler bærer en uendelig omkostning, anses lokaler, der mangler faciliteter efterspurgt i bookingsituationen ("*nødvendige faciliteter*") for uegnede og tildeles en prohibitivt høj omkostning.

Omvendt justeres omkostningerne for lokaler, der besidder alle eller nogle af de opregnede nødvendige faciliteter, baseret på, hvor godt de matcher nødvendige faciliteter. Jo tættere matchet er, desto lavere er omkostningen. Vi går ud fra, at et lokales faciliteter ikke ændrer sig over tid. Et lokale betragtes som brugbart, hvis det indeholder alle de faciliteter, der er angivet som nødvendige. For disse egnede lokaler beregnes omkostningerne baseret på forskellen mellem antallet af faciliteter i lokalet og antallet af nødvendige faciliteter for aktiviteten.

Denne fremgangsmåde afspejler kapacitetsomkostningsmodellen, hvor omkostningsfunktionen påvirkes af det relative match eller mismatch mellem, hvad der er nødvendigt, og hvad der er tilgængeligt. Lokaler med præcis de nødvendige faciliteter ville ikke medføre nogen ekstra omkostning, mens lokaler med flere faciliteter end nødvendigt ville medføre en omkostning proportional med overskuddet. Dette sikrer, at lokaler ikke kun vælges for deres fysiske kapacitet, men også for deres funktionelle egnethed, tæt afstemt med specifikationerne for den planlagte aktivitet.

7.7.3 anbefalinger

Den samlede omkostningsfunktion baseret på fire isolerede omkostningsmodeller, anvendes praktisk i et anbefalingssystem. Ved projektets afslutning er mediet eller metoden for hvordan en bruger bør interagere med EAB endnu ikke fastlagt. Arkitektonisk vil EAB f.eks. kunne eksponeres som en mikroservice i en webapplikation lig det setup der anvendes for Driftoptimeringsmodellen. Givet EABs eksperimentelle stadie, er en fuld løsningsarkitektur endnu ikke relevant. Figur 41, nedenfor, viser et *user flow*: Hvordan en bruger vil kunne interagere med systemet. Hvorvidt EAB bør fungere som selvstændig løsning eller som komponent i et eksisterende bookingsystem er endnu ikke besluttet.



Figur 41: User flow for EAB

Til projektets formal er en simpel webapplikation udviklet som demonstration af hvordan EAB kunne eksponeres til brugere som en selvstændig service. Hvordan denne applikation spindes op er beskrevet i den tekniske dokumentation for projektet.

7.7.4 anbefalinger i realtid

EAB er udviklet mod at kunne håndtere bookingsituationer i realtid. Fordi at vurderingen af hvilken bookingmuligheder der er at foretrække, er kontekstafhængig er dette en præmis for løsningen. I en bookingsituation vil anbefalingssystemet modtage specifikationer (Antallet af personer, nødvendige faciliteter osv.) fra en bruger og sende disse informationer videre til omkostningsmodellerne. Her antages det, at omkostningsmodellerne har adgang til

- En oversigt over hvilket faciliteter der er til rådighed i hvert lokale
- En oversigt over kapaciteten for hvert lokale
- En komplet opdateret bookinghistorik op til tidspunktet for denne bookingsituation.

Bemærk at energikoefficienter udregnet ovenfor behøves at blive regnet for hver bookingsituationer med i stedet forbliver statistiske i en længere periode før en eventuel gentræning indtræffer. De ovenstående faktorer er nødvendige for at sætte EAB i produktion, og har været i projektet være en præmis for de bagvedliggende eksperimenter.

Når omkostningsmodellerne har returneret deres foreslåede omkostninger, summeres disse for hvert lokale. Derefter rangeres lokalerne efter denne summerede omkostning, således at de "*billigste*" lokaler præsenteres som de bedste bookingmuligheder til brugeren.

7.8 Evaluering og perspektiver

Det energibevidste AI bookingsystem præsenteret i denne sektion anvender en blanding af datadrevne og regelbaseret omkostningsmodeller til at vurdere bookingmuligheder i realtid. Vores indledende integrationstests og modevaluering (som i figur 40 ovenfor) viser lovende resultater, men det bør understreges at EAB ikke er en produktionsfærdig løsning. Det kræver en efterfølgende faste af videreudvikling for at bringe løsningen til et gunstigt stadie. Mere grundlæggende er dog datagrundlaget: I de ovenstående omkostningsmodeller og de bagvedliggende eksperimenter, introduceres en række simplificerende antagelser, hvor mange er relateret til mangler i projektets datagrundlag. F.eks. er alle informationer om kapacitet og faciliteter i hvert lokale syntetiske. Ligeledes er det vigtigt at understrege at vores modeldomæne, dvs. en approksimation af virkeligheden som vi kan repræsentere matematisk og igennem data, er simpel i forhold til virkeligheden for energioptimering i en bygning.

De udførte beregninger på energiforbrug giver positive resultater under simplificerende antagelser. Indenfor et konfidensinterval af +/- 7 kW/h er vi i stand til at forudsige op imod 95% af lokalers energiforbrug på udvalgte skoler og de afledte energikoefficienter er et bærende element i den samlede omkostningsfunktion bag EAB. Dette bør tjene som et fundament for refleksion over videreudvikling: Hvad kan opnås hvis energiomkostninger der isoleres til et givent lokales aktivitet? Den viden kan, udover at bidrage til en energibevidst bookingproces, også finde anvendelse i fremtidig energianalyse.

Givet det abrupte retningsskift for AI Signaturprojektet er en fyldestgørende evalueringsfase ej heller blevet realiseret. Det betyder at, selvom vores interne evalueringer og resultater af energimodellering er lovende, har vi ikke en robust grundlag for den samlede løsnings performance. Havde vi fortsat udviklingsarbejdet igennem hele projektet ville den opnåede evalueringsproces have givet et mere håndgribeligt overblik over hvilke styrker og svagheder det valgte design har.

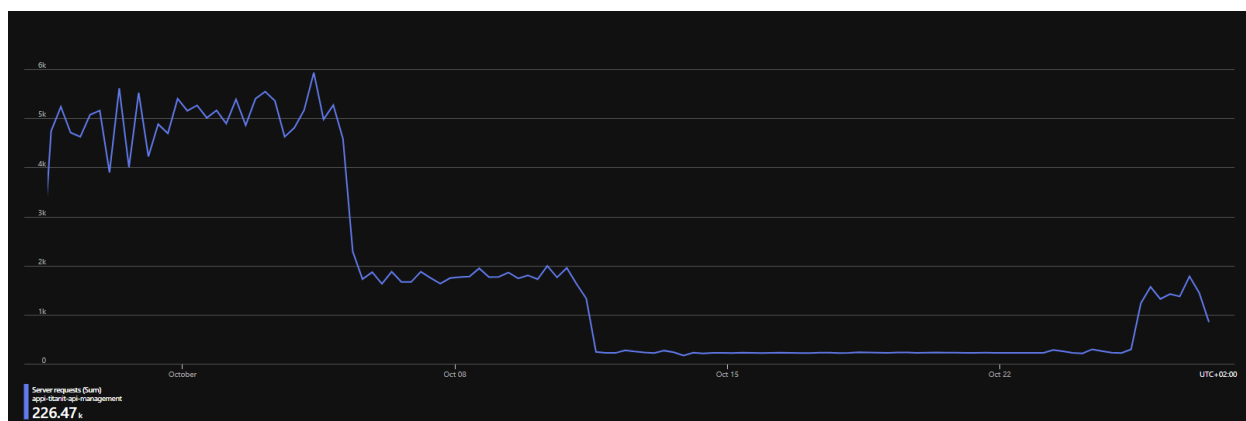
7.9 Konklusion

Samlet set er EAB en velstruktureret men ufærdig løsning med potentiale for at introducere signifikante besparelser samt automatisere visse dele af bookingprocessen. Givet afslutningen på processen, er der rig mulighed for at videreudvikle konceptet frem mod en *proof-of-concept*-løsning der kan AB-testes eller evalueres mod tidligere års energiforbrug. De initiale energiberegninger viser lovende resultater og designet tillader tilføjelsen af nye omkostningsmodeller. På den måde er EAB yderst skalérbar på sit nuværende stadie. Der er endnu uafklaret hvorledes EAB skal eksponeres for brugere af systemet, men det bagvedliggende design vil kunne fungere både som selvstændig webapplikation og integreret komponent i et eksisterende bookingsystem.

8. Generelle udfordringer med data

Allerede i projektets indledning blev processen med dataindsamling startet. Målet var at indhente, i første omgang, historisk data fra IoT-sensorer, Booking-systemer og energi-systemer samt indsamle stamdata på skolens lokaler, arealer m.m. Processen tog dog længere tid end først antaget og der har i hele projektperioden løbende kommet ny data til. Derudover har der også været udfordringer med tilgængelighed af data, da ikke alle 3 udvalgte test-skoler har haft samme datagrundlag. Der er derfor forsøgt at arbejde med det data der går igen for alle 3 skoler. Foruden den historiske data har projektet også skulle bruge "live" adgang til data via API'er. Disse har også været igennem en længere proces for at få indhentet. Alt i alt tog dataindsamlingen komplet 8-9 måneder i stedet for den først antaget ene måned. Generelt har arbejdet med ensretning ift. datoer, formater og kvalitet også været en større opgave. Alle udfordringerne er forsøgt løst ved at skabe tabeller med ensrettet formater som kan være input til det fremtidige arbejde med udkommet af projektet.

Det største issue med data har været omkring Zurface-enhederne, der er installeret i Aarhus og Favrskov til Indeklima målinger. Her er der integreret med platformens eksisterende API ved hjælp af OS2-IoT. Dog har datafrekvensen været inkonsistent og antallet af forsendelse af pakker har været svingende. Dette fremgår også af nedenstående:



9. Fremtidige perspektiver

Projektet har efter sin afslutning opstillet flere perspektiver og områder som vil være værdifulde at arbejde videre med for at få et reelt produkt ud af AI-modellerne og dashboards.

9.1 Analytics

Fra et dataperspektiv vil det være enormt gavnligt at kommunerne individuelt arbejder med at ensrette data internt til at passe ind i de foreslåede skemaer. Disse skemaer kan ses i GitHub repository Enformanten under docs/analytics/transformations/general og under base table schemas. Dette gøres i form af et "context-broker" system, der kan sikre at pakker med energi og IoT data indeholder den fornødne information som standard. Dette kan f.eks. være at en datapakke med indeklimate data også indeholder lokale ID'er/skole i stedet for blot et device ID. Alt i alt vil det gøre arbejdet med forskellige datakilder mere strømlinet og behovet for individuelle transformationer vil ligge hos kommunerne selv i stedet for i et centraliseret værktøj.

Datamodellen kører i dag på et rent kommune-begreb, der gør det muligt at se data fra én skole per kommune. Derfor vil det være naturligt at få udvidet modellen til at kunne indeholde segrerret data på både kommune og bygningsniveau.

Datagrundlaget generelt vil også være fordelagtigt at solidere ved at opsætte flere sensorer med konsistent datagrundlag. I projektet har der været indsamlet data på luftkvalitet og bevægelse foruden CO2 som kunne være en idé at kigge ind i forhold til Driftoptimeringsmodellen. Det vil derfor også være en idé at få sat sensorer op i samtlige relevante lokaler for en mere præcis modellering for en hel bygning. Det sidste forslag til fremtidige projekter er at få håndteret behovet for manuelle benchmark der knytter sig til enkelte kommuner eller bygninger. Dette kræver et setup hvor kommunerne generelt kan styre dashboards og udregninger individuelt.

9.2 Driftoptimeringsmodellen

Driftoptimeringsmodellen er per leveringsdato produktionsklar som selvstændigt komponent i Enformantens overordnede arkitektur. Men givet det relative korte design- og udviklingsforløb har det langt fra været muligt at bringe løsningen til dens fulde potentiale. I det følgende gennemgås en række vigtige overvejelser vedrørende DOMs nuværende design og mål for videreudvikling.

9.2.1 Evaluering og annoteret data

Den primære iboende udfordring ved det nuværende design er evalueringsmetoden. Som uddybet (flere gange) ovenfor, har vi ikke adgang til pålidelig annoteret data, hvilket komplicerer anvendelsen af traditionelle metrikker til evaluering af statistiske modeller, såsom *F1-score*, *precision* og *recall*. Omvendt beror de fundamentale designvalg netop på fraværet af annoteret data. I et videre udviklingsforløb hvor annoteret data kan tilføjes datagrundlaget, bør modelleringstrategien genovervejes. En superviseret løsning vil givetvis kunne matche eller overgå den performance vi opnår igennem det eksisterende design. Mens annoteret data ikke er tilgængeligt eller nemt at fremskaffe på kort sigt, vil der givetvis være muligt at annotere data for en række lokaler manuelt, endda assisteret af de eksisterende modeller. Denne manuelt annoteret data kunne derefter anvendes som et såkaldt *gold-labelled test set*, dvs. et datasæt fri for fejl som kan bruges til at evaluere performance.

9.2.2 Dataintegritet

Uden at forbedre integriteten af de anvendte data, vil enhver videreudvikling rettet mod performanceoptimering være overflødig. Ved f.eks. at sikre at deltagende skoler og kommuner kan levere konsistente data fra bevægelsessensorer, vil vi kunne udvide omfanget af brugbare data, hvilket givetvis vil resultere i bedre performance. I skrivende stund er AI-modellerne begrænset til at udlede brugsstatus for et givent rum baseret på kun en håndfuld variabler, med CO2 som det bærende element, grundet det inkonsistente datagrundlag. Hvis datakvaliteten for andre sensoriske variabler, så som temperatur, IAQ og bevægelse kan bringes op på samme niveau, vil det givetvis have en positiv effekt på AI-modellernes performance. Ligeledes forholder det sig med den datastrøm som Driftsoptimeringsmodellen modtager løbende: Hvis der i denne data opstår huller og/eller fejlmeldinger, vil det have en negativ effekt på modellerne, hvis ikke forbehandlingsmodulet optimeres til at adressere netop disse udfordringer sideløbende. Herudover vil indførelsen af nye sensoriske variabler og/eller metadata så som specifikationer af antal vinduer, kapacitet og andre karakteristikker per rum givetvis kunne influere AI-modellernes performance positivt.

9.2.3 Onboarding af nye bygninger

Arkitekturen er fortrinsvis agnostisk i forhold til antallet af bygninger og lokaler der tilføjes, idet modelregisteret gemmer rumspecifikke modeller under et unikt ID udledt af skolens (eller bygningens) navn og lokalets ID. Løsningen har vi sig kampdygtig op imod 1.8 mio. tidsintervaller over 164 rum (håndteret på omkring 5 minutter), men ved yderligere skalering bør der tages højde for den per-lokale komputationelle hastighed i måden forbehandling orkestreres på. Stresstest af systemet bør prioriteres ved opskalering.

9.3 Energibevidst AI Booking

EAB er per leveringsdato ikke produktionsklar som selvstændigt komponent i Enformantens overordnede arkitektur. I det følgende gennemgås en række vigtige overvejelser vedrørende EABs nuværende design og mål for videreudvikling.

9.3.1 Dataintegritet

At forbedre datagrundlaget for et EAB vil blandt andet kræve en overgang fra syntetiske til mere realistiske data. Dette indebærer indsamling af faktiske brugs- og specifikationsdata fra relevante lokaler. Vi ved fra analysen bag Driftoptimeringsmodellen at en booking ikke altid er lig med aktivitet. Derfor er fundamentet for f.eks. energiberegningerne ikke retfærdigt: En model trænet på kureret aktivitetsdata vil givetvis være bedre i stand til at afdække sammenhænge i relation til energiforbrug. Ved at indsamle data om faktisk aktivitet og lokaletspecifikationer, så som kapacitet, faciliteter og brugsmønstre, kan EAB udbygges til et virkelighedsnært, robust værktøj til at understøtte bookingprocessen. Denne proces vil også give mulighed for at identificere og korrigere for eventuelle skævheder eller unøjagtigheder i de oprindelige antagelser. Grundlæggende vil en komplet repræsentation af dette domæne være et enormt foretagende. Givetvis vil vi dog, ved blot at tilføje datagrundlaget få flere datakilder af høj integritet, kunne bygge et langt mere repræsentativt system.

9.3.2 Tilpasning og videreudvikling af omkostningsmodeller

Udviklingen af et mere repræsentativt modeldomæne indebærer at introducere flere omkostningsmodeller til at beskrive komplekse interaktioner, der påvirker energiforbrug i bygninger. Dette kunne omfatte faktorer som bygningens alder, isolering, vejrforhold, og brugernes adfærd, inklusive brugsmønstre i skoletiden. Ved at udvide antallet af omkostningsmodeller bevæger vi os mod en mere komplet repræsentation hvilket, alt andet lige, vil have en gunstig effekt på systemets performance evne til at give korrekte energibevidste anbefalinger.

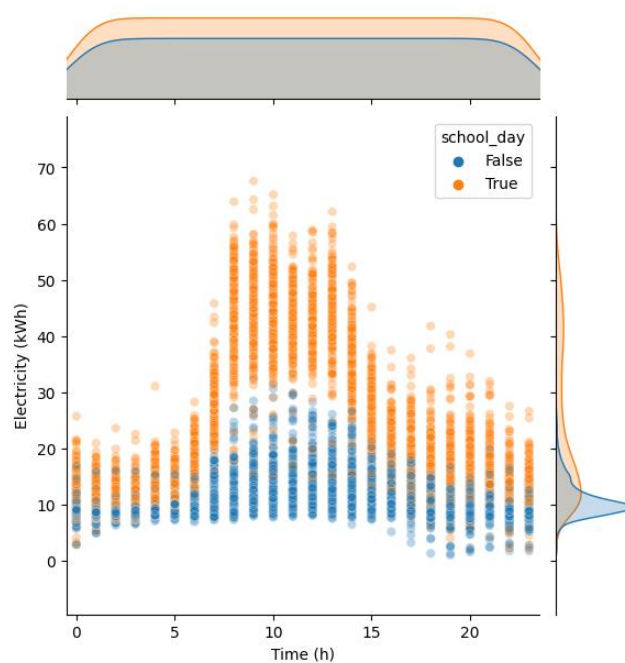
9.3.3 Optimering af energiberegninger

Forbedring af energiberegninger kræver kan både bero på en opdatering af datakvaliteten og en mere dedikeret ingeniørgerning i modelleringsregi. En oplagt videreudvikling er introduktionen af tidsrækkeanalyser indenfor brugsadfærd for at afdække chancen for at et lokale vil blive taget i brug, både før og efter en booking er placeret.

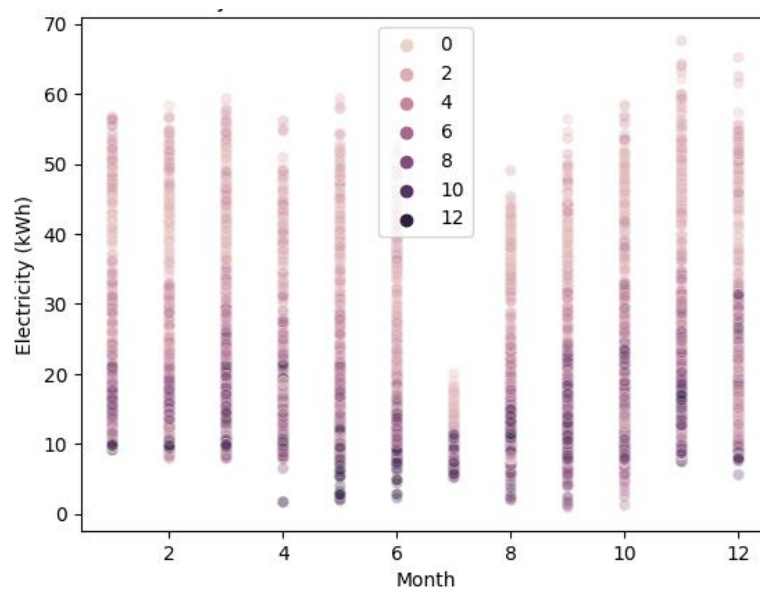
9.3.4 Evaluering

En grundig evaluerings- og testfase for EAB vil gøre os i stand til at estimere den samlede besparelse som en serie af energibevidste bookinger vil kunne introducere relativt til en syntetisk stokastisk process eller et faktisk bookingmønster. Her vil evaluering af slutbrugere være et vigtigt element. Udfordringer ved evaluering af EAB er beslægtet med udfordringen for Driftoptimeringsmodellen: Selv hvis vi kan anbefale fortrinsvis mere energibevidste bookingmuligheder, hvad evaluerer vi så performance imod? Et brugsorienteret system som EAB vil kunne introduceres i en AB-test for at evaluere hvorvidt systemet fører til en besparelse. Alternativt vil systemet kunne evaluere imod tidligere års booking- og aktivitetshistorik.

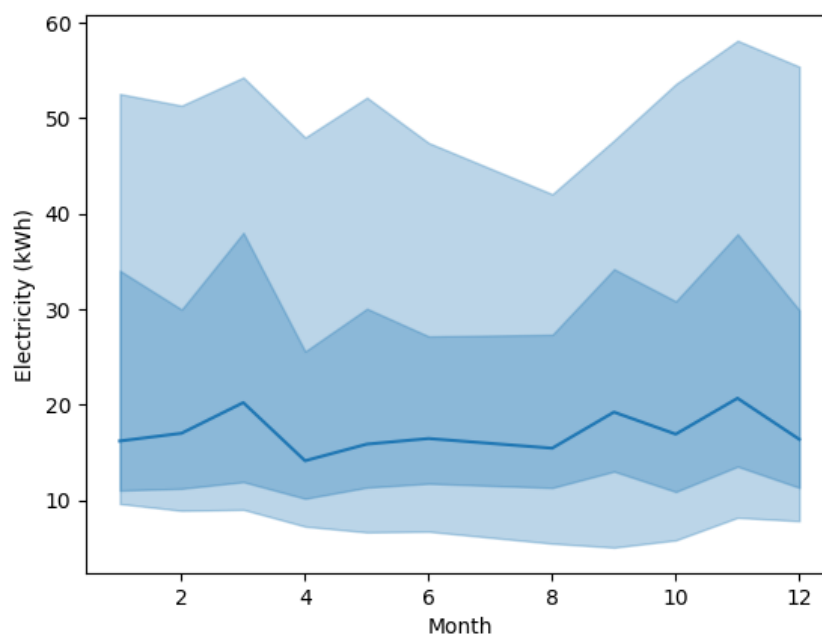
10. Appendix A



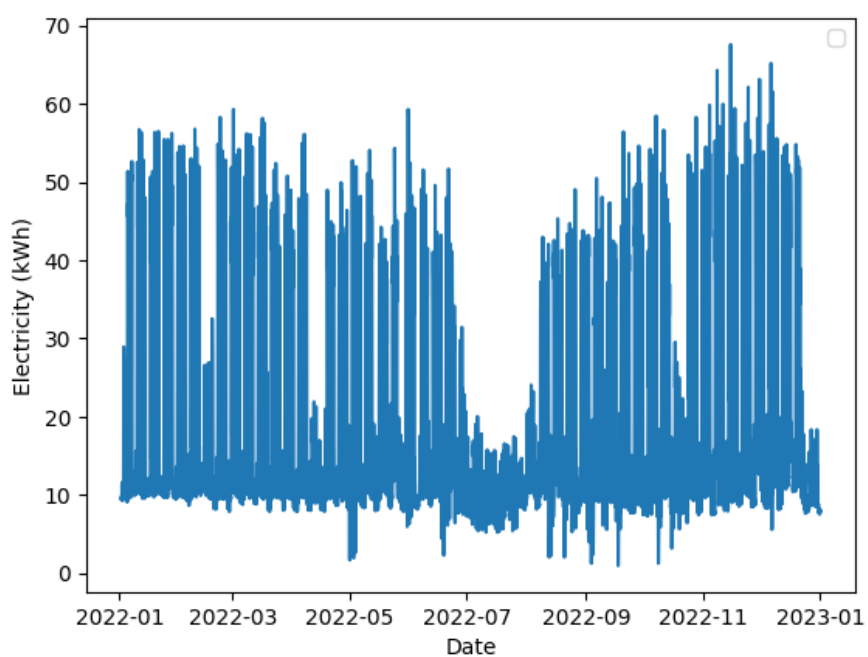
Figur 39: Energiforbrug per time, opdelt skoledage og ikke-skoledage (weekender, ferie mv.) (Strandskolen, Aarhus Kommune)



Figur 40: Energiforbrug per dagtime, per måned (Strandskolen, Aarhus Kommune)



Figur 41: Minimum og maksimum energiforbrug per måned (Strandskolen, Aarhus Kommune)



Figur 42: Energiforbrug over tid (Strandskolen, Aarhus Kommune)