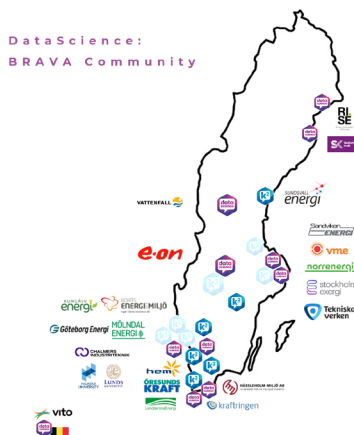
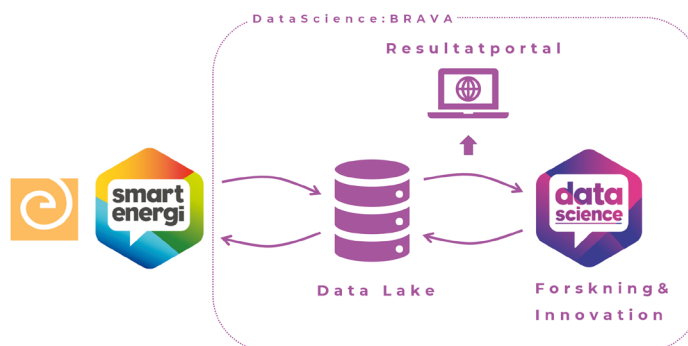
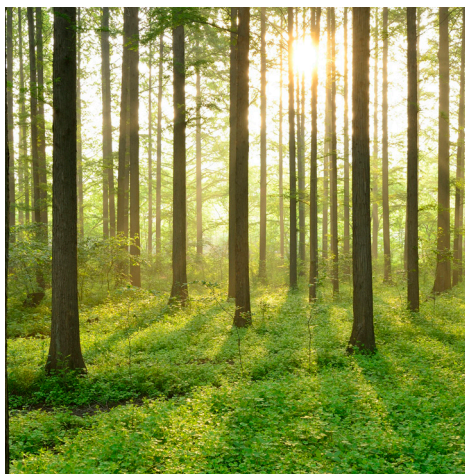


BRANSCHSAMARBETE FÖR AVANCERAD ANALYS AV VÄRMEDISTRIBUTION OCH UPPVÄRMNINGSBEHOV

RAPPORT 2022:850



Branschsamarbete för avancerad analys av värmedistribution och uppvärmningsbehov

Data Science DS:BRAVA – Ett projekt för smartare
fjärrvärme

**FILIP WÄSTBERG
MARIA HANSSON
RIKARD EDLAND**

ISBN 978-91-7673-850-4 | © Energiforsk febr 2022

Energiforsk AB | Telefon: 08-677 25 30 | E-post: kontakt@energiforsk.se | www.energiforsk.se

Förord

Det här projektet startade med syfte att öka den analytiska förmågan i fjärrvärmebranschen. Projektets har dels svarat på olika analysutmaningar som finns i branschen och hur branschen kan samarbete kring digital innovation och vilka förutsättningar som krävs.

Projektet har letts av Maria Hansson från Solita och genomförts tillsammans med kollegan Filip Wästberg och Rikard Edland från Chalmers tekniska högskola. En referensgrupp bestående av Fabian Levhin, Stockholm Exergi (sammanhållande); Lisa Smitt, Skellefteå Kraft; Andreas Carlsson, Borås Energi och Miljö; Jessica Englund, E.ON Energiinfrastruktur; Johan Grünwedl, Norrenergi; Tobias Rehnholm, Vattenfall; Thomas Franzén, Göteborg Energi; Henrik Akselsson, Tekniska verken i Linköping och Fredrik Nyman, Krafringen har följt och kvalitetssäkrat projektet.

Projektet ingår i programmet FutureHeat vars långsiktiga mål är att bidra till visionen om ett hållbart uppvärmningssystem med framgångsrika företag som utnyttjar nya tekniska möjligheter och där de samhällsinvesteringar som gjorts i fjärrvärme- och fjärrkyla tas till vara på bästa sätt. Detta projekt ingår i programmets andra etapp.

Programmet leds av en styrgrupp bestående av Jonas Cognell, Göteborg Energi (ordförande); Anders Moritz, Tekniska verken i Linköping; Anna Hinderson, Vattenfall AB; Charlotte Tengborg, E.ON Energiinfrastruktur; Fabian Levhin, Stockholm Exergi; Holger Feurstein, Krafringen; Dan Bruhn, Jönköping Energi; Patrik Grönbeck, Borlänge Energi; Leif Bodinson, Söderenergi; Lena Olsson Ingvarson, Mölndal Energi; Magnus Ohlsson, Öresundskraft; Niklas Lindmark, Gävle Energi; Per Örvind, Eskilstuna Strängnäs Energi & Miljö; Petra Nilsson, Växjö Energi; Staffan Stymne, Norrenergi; Stefan Hjærtstam, Borås Energi och Miljö; Svante Carlsson, Skellefteå Kraft; Ulf Lindquist, Jämtkraft och Julia Kuylenstierna (adjungerande), Energiforsk. Suppleanter utgörs av Ann Britt Larsson, Tekniska verken i Linköping; och Peter Rosenkvist, Gävle Energi.

Julia Kuylenstierna, programansvarig FutureHeat

Här redovisas resultat och slutsatser från ett projekt inom ett forskningsprogram som drivs av Energiforsk. Det är rapportförfattaren/-författarna som ansvarar för innehållet.

Sammanfattning

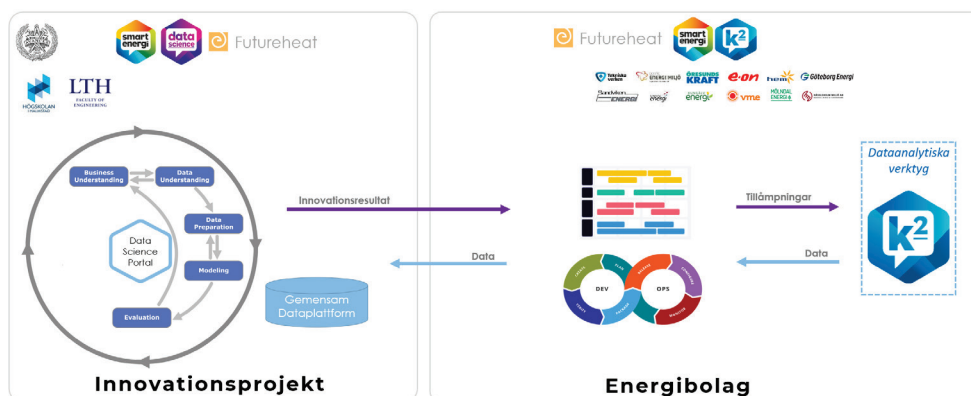
Genom ett omfattande samarbete mellan fjärrvärmebolag, akademi och data-science- expertis har innovations- och forskningsprojektet DS:BRAVA banat vägen för framtidens smarta fjärrvärme. Syftet är att visa på hur man kan etablera en effektivare och snabbare cykel för att arbeta fram nya och ständigt förbättrade metoder inom avancerad analys- och maskininläring/AI för att bli bättre på att hitta avvikelser och förbättringsmöjligheter inom hela värmedistributionen.

Projektet har visat på samverkan inom tre delar av data-science-cykeln; att samverka inom branschen för att harmonisera och samla in FAIR data (findable, available, reusable, interoperable) från sveriges fjärrvärmecentraler, att samverka mellan energibolag och universitet kring avancerad analys och AI och att samverka i utrullning av tillämpad intelligens, dvs. att tillämpa resultat i mjukvara för daglig användning hos energibolagen.

Cirkeln har slutits: Inom ramen för projektet har en taxonomi för etikettering av data tagits fram. Taxonomin skapar stringens för branschen och gör det enklare för företag att märka upp fel-åtgärder-orsaker i fjärrvärmedata för att kunna träna maskininlärningsmodeller. Fjärrvärmedata har samlats från energiföretag, framförallt från analysapplikationen K2 till en gemensam dataplattform kopplad till en data science-portal.

I och med att projektet dessutom har kunnat validera de analysmodeller och ramverk som projektet tagit fram i den branschgemensamma applikationen K2, så har cirkeln slutits; från tillämpning delas data, från data till analys, från analys till innovation, från innovation till tillämpning.

I längden kommer detta leda till en bättre monitorerad och mer optimerad värmedistribution.



*DS:BRAVA är förkortning för projektnamnet Branschgemensam Avancerad Analys.

Nyckelord

AI, avvikelser, avvikelседetektering, taxonomi, smart fjärrvärme, smart energi, machine learning, felkoder, data science, proaktivt underhåll, värmedistribution, labeled data, digital infrastructure

Summary

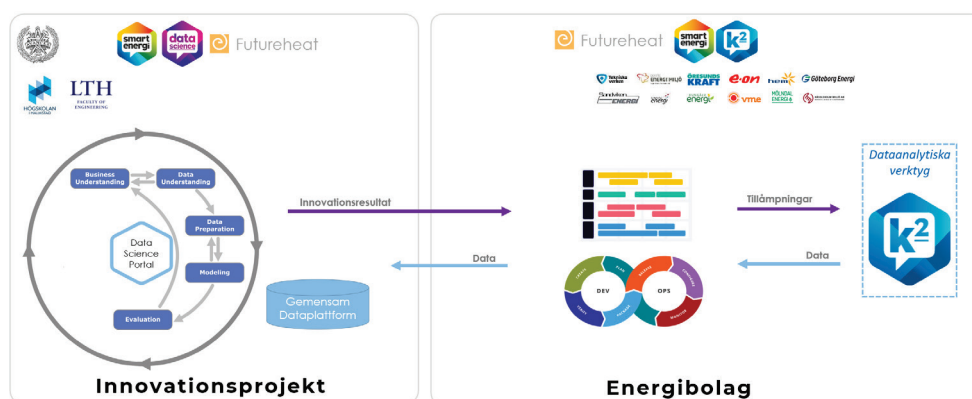
Through extensive collaboration between district heating companies, academia and data science expertise, the innovation and research project DS: BRAVA has paved the way for the smart district heating of the future. The purpose is to show how to establish a more efficient and faster cycle to develop new and constantly improved methods in advanced data analytics and machine learning / AI to become better at finding deviations and improvement opportunities within the entire heat distribution.

The project highlights the benefits and need of collaboration in three areas of the data science lifecycle; 1) industry collaboration to harmonize and collect FAIR data (findable, available, reusable, interoperable) from Sweden's district heating substations 2) collaborations between energy companies and universities in advanced data analysis and AI 3) collaborate in the rollout of applied intelligence, i.e., applying results in software for daily use in the industry.

Full circle: Within the scope of this project, a taxonomy for labeling data has been developed. The taxonomy creates stringency for the industry, making it easier for companies to mark fault-measure-cause in district heating data to be able to train machine learning models. District heating data has been collected from energy companies, primarily from the analysis application K2 to a common data platform connected to a data science portal.

As the project has also been able to validate analytical models and frameworks using the industry-wide application K2 the data science lifecycle-circle has been closed; data from software applications is shared, data is used to perform analytical reasoning, results from analysis are described in innovative methods and frameworks, innovation is ready to be applied in software application.

In the long run, this will lead to a better monitored and more optimized heat- and cooling distribution.



* DS: BRAVA is an abbreviation for the project name Branschgemensam Advanced Advanced Analysis.

Innehåll

1	Inledning och bakgrund	9
1.1	Avsaknad av data med hög kvalitet	9
1.2	Avsaknad av kompetens och verktyg	9
1.3	Mål och syfte	10
1.4	Finansiering och projektorganisation	10
2	DS:BRAVA Data Science-portal	13
3	DS:BRAVA Data Lake	15
4	Avvikelseorsakstaxonomi	16
4.1	De fem nivåerna i avvikelseorsakstaxonomi	17
4.1.1	Avvikelseorsak	17
4.1.2	Komponent	18
4.1.3	Felbeskrivning	18
4.1.4	Åtgärd	18
4.1.5	Status	19
4.2	Detaljerade beskrivningar av komponentnivåerna	19
4.3	Implementering av avvikelseorsakstaxonomi	22
5	Utveckling av analysmodeller	25
5.1	Arnold	27
5.2	Referensperiod	28
5.3	Identifiering av läckage i servisledning med hjälp av Förändringspunktsanalys	28
5.4	Virtuella grannar	29
5.5	Modeller Halmstad Universitet	30
5.6	Analys av energideklarationer	30
5.7	Multipla energisignaturer	31
5.8	Extremvärden och identifikation av anomalier	32
6	Implementering av modeller i mjukvara	33
6.1	Ramverk för stringens i mjukvarutillämpning: DS:BRVALearn	33
7	Analysvärdesmatris för avancerad analys	34
8	Community, kunskapsdelning och publika webinar/event	36
9	Utblick för fortsatta aktiviteter	38
9.1	Förslag på uppföljningsaktiviteter	38
9.1.1	Förvaltning av Data Science Portal och Data Lake	39
9.1.2	Kompetenshöjande aktiviteter i branschen	39
9.1.3	Datadelning för forskningssyfte	39
9.1.4	Förvaltning av taxonomin	39
9.1.5	Analysvärdesmatrisen och dashboard	39
9.1.6	Arbetsflöde för identifiering av avvikelser	39

9.1.7	Tillämpning av avancerad analytiska modeller	39
9.1.8	Identifiering av läckage + testbädd	40
9.1.9	Koppling mellan energibolag och fastighetsbolag	40
9.2	Potentiella offentliga finansieringsmedel för fortsatta steg	41
10	Referenslista	42

1 Inledning och bakgrund

Enligt rapporten *Fault detection in district heating substations* [5] konstateras att **75% av alla fjärrvärmecentraler innehåller fel som skulle kunna upptäckas genom att analysera avvikelser**. Avvikelseanalysen kan ske manuellt men för att kunna skala och erbjuda konkurrenskraftiga metoder behöver avvikelседetekteringen automatiseras och **dataanalytiska metoder är då ett mycket effektivt angreppssätt** för detta.

Att identifiera avvikelser i data från fjärrvärme är svårt och tar tid. De flesta fjärrvärmebolag har initiativ för att med avancerad analys bli bättre på att identifiera avvikelser och fel i fjärrvärmesystem. Problemen som företagen stöter på är snarlika. Det saknas kompetens, skalbara verktyg men framför allt de mängd data som behövs, **branschstringent eller standardiserat märkt data som kan användas både i gemensamma verktyg och i forsknings- och innovationsprojekt** så som detta projekt.

I och med att fjärrvärmedistributörer står inför samma eller liknande utmaningar kan innovation och lösningar med stora fördelar delas. Dessutom kan branschen tillsammans höja kompetensen och ta gemensamt ansvar för den utveckling som behövs.

Det finns dessutom stora konkurrensfördelaktiga och **ekonomiska värden av att proaktivt arbeta med avvikelse- och åtgärdsplanering**. En utmaning för fjärrvärmebolagen är att behålla kunder, eftersom en tappad kund innebär ett inkomstbortfall som ej går att jämföra med den konstadsminskning det innebär av förlorad kund. Fjärrvärmeproduktion och distribution behöver arbeta med proaktiv optimering för att vårda sina kundrelationer, stärka konkurrensfördelar och motivera den samverkanspotential fjärrvärme har i sin grundidé.

1.1 AVSAKNAD AV DATA MED HÖG KVALITET

De statistiska modellerna, brett kallat för AI, ligger bakom exempelvis självkörande bilar och ansiktsigenkänning. De är i regel utvecklade på data med tydliga etiketter (eng. *labeled data*). Om vi först låter en människa märka upp exempelvis bilder på katter och hundar kan vi sedan dela upp bilderna i pixlar, omvandla dessa till siffror och träna en algoritm eller statistisk modell som kan replikera människans sortering. Det är vad vi brukar kalla för maskininlärning.

I stort sett alla fjärrvärmebolag har initiativ för att bättre kunna identifiera och hitta avvikelser i data. Ett av de största problemen är avsaknaden av branschstringent märkt/etiketterad data. Därför har projektet DS:BRAVA lagt en stor del av tiden på att utveckla en taxonomi och implementering av felkoder som kan användas för att göra det enklare att etikettera data.

1.2 AVSAKNAD AV KOMPETENS OCH VERKTYG

Det går att komma framåt även utan märkt data/etiketter. Branschen behöver även arbeta med att sprida kunskapen om dataanalytiska metoder, vad det innebär,

ställer för krav, vilka möjligheter det ger, samarbetsfördelar osv. Branschen behöver höja kompetensen i allt ifrån kundnära/anläggningsnära roller, till verksamhetsutveckling, planering, prioritering och ledning. DS:BRAVA har arrangerat flertalet publika webinar och event där resultat presenterats kontinuerligt, kunskap har delats och deltagare kunnat diskutera och ställt frågor.

Att dessutom främja samarbete mellan universitet, företag och industrispecialister är också av stor vikt. DS:BRAVA-projektet har haft som fokus på att göra det enkelt för forskare och specialister från olika branscher att dela med sig av kod, resultat och data. Det har framför allt skett genom en gemensam Data Lake och en gemensam Data Science-portal.

För att snabbare komma framåt i avancerad analys inom fjärrvärme, från idé till implementering, behövs gemensamma ramverk för kod. Företag som säljer analystjänster delar av förklarliga skäl inte med sig av koden. Projektet DS:BRAVA fokuserar därför på att smidigt kunna sprida kod och analysresultat mellan samarbetsparter samt erbjuda en digital infrastruktur för att möjliggöra detta ramverk i praktiken. Smart Energi tillhandahåller data från energibolagen genom att dela data från applikationen K2 till DS:BRAVA Data Lake. Datadelningen sker i form av ett Datadelnings-API som Energibolagen själva kan aktivera. Data delas därefter anonymiserat från respektive energibolags applikation K2 till DS:BRAS Data Lake. DS:BRAVA tar del av datat och bedriver forskning- och innovation på denna. Resultat i projektet kan sedan utvärderas och valideras i mjukvara genom att vidare undersöka hur innovationsresultat skulle implementeras i praktiken, dvs. i mjukvara som energibolagen kan ta del utav. Cirkeln sluts.

1.3 MÅL OCH SYFTE

Målet för projektet DS:BRAVA är att utifrån energibolagens prioritering ta fram verifierbara analysmodeller som kan användas för att identifiera läckage, styrfel och energimönster. För att lyckas med det här behövs en plattform där energianalytiker, verksamhetsutvecklare, mjukvaruutvecklare och data scientists kan ta del av samlad tränings- och testdata samt analysmodeller.

För att möjliggöra mer avancerad dataanalys hos energibolagen, i form av avancerad dataanalytiska och AI-modeller behövs etiketterad data. Därför behövs också ett bra sätt för företag att själva kunna etikettera data.

Därtill behövs en enkel metod för företag och organisationer att kunna identifiera värdet av avancerad analys.

1.4 FINANSIERING OCH PROJEKTORGANISATION

Projektet är ett samarbete mellan Sveriges Energibolag i form av branschsamarbetsorganisationen Smart Energi, Eneriforsk:Futureheat, Energimyndigheten, Halmstad Högskola, Chalmers Industriteknik. Även Lunds Tekniska Högskola, Borås Högskola och RISE har haft stor betydelse för projektet och dess resultat. Solita har haft projektuppföljning- och koordineringsansvar, samt bidragit med expertresurser inom Data Science/AI/ML, Data Management och mjukvaruutveckling. Smart Energi har tillhandahållit den tekniska

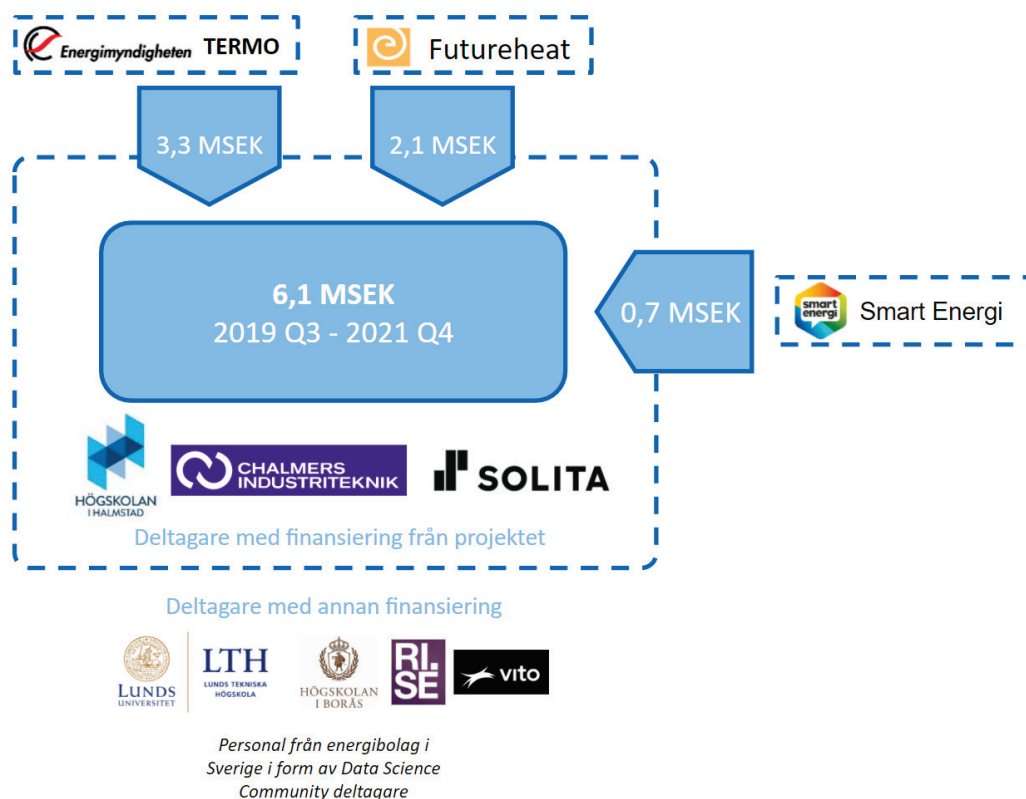
infrastrukturen för projektet i form utav tekniska miljöer för Data Lake, Data Science Portal (Kyso), samt automatiska datadelningsmekanismer från Smart Energi Energibolags applikation K2.

Samtliga medlemsbolag i Futureheat och Smart Energi har rätt till projektresultaten och kan själva välja hur de ska tas vidare till tillämpning. T.ex. kan enskilt bolag beställa tillämpning av resultat i valfri mjukvara eller arbetsprocess, alt. kan flera bolag samverka kring tillämpning och sambruka det i gemensam/delad mjukvara. Projektet har inte tagit fram rekommendationer eller instruktioner för hur resultat ska tillämpas.

Ett Data Science-råd med representanter från både energibolag, expertis inom såväl fjärrvärme som data science har kontinuerligt bidragit med resonemang kring prioriteringar och förväntat värde av olika moment i projektet.

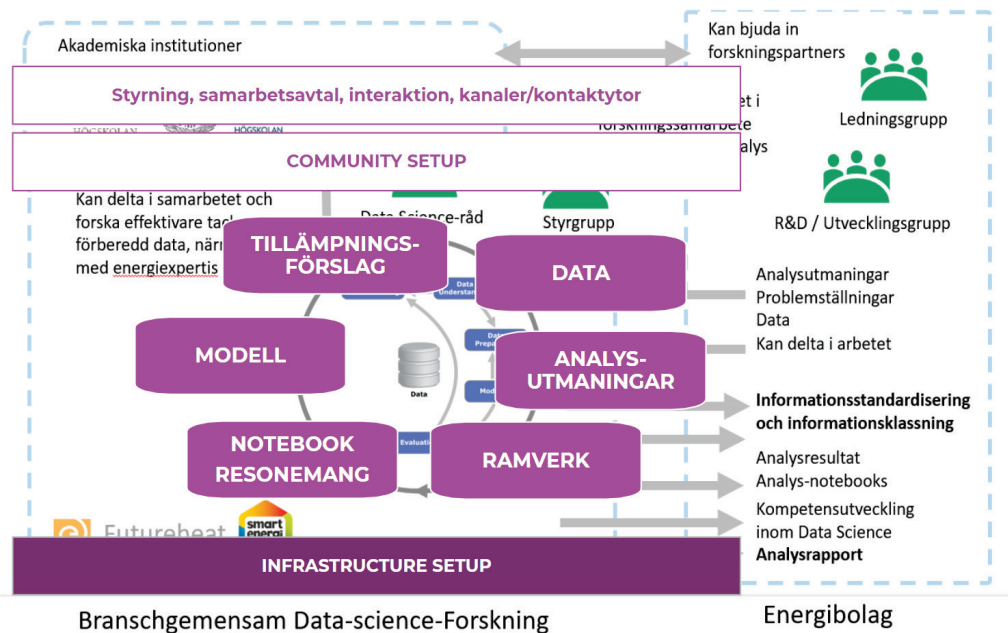
Projektet är finansierat av Energimyndigheten TERMO, Energiforsk Futureheat och Smart Energi samarbetsorganisation. Bilden nedan, Figur 1, förklarar finansieringen och hur deltagare i projektet har finansierats.

Samtliga personer som deltagit i projektets publika webinar och diskussioner har bidragit till projektets resultat och måluppfyllnad.



Figur 1 Schematisk bild över projektets finansiering

Figur 2 nedan relaterar detta branschgemensamma innovationsprojekt DS:BRAVA till respektive Energibolags egna deltagande, under projektet eller efter projektet.

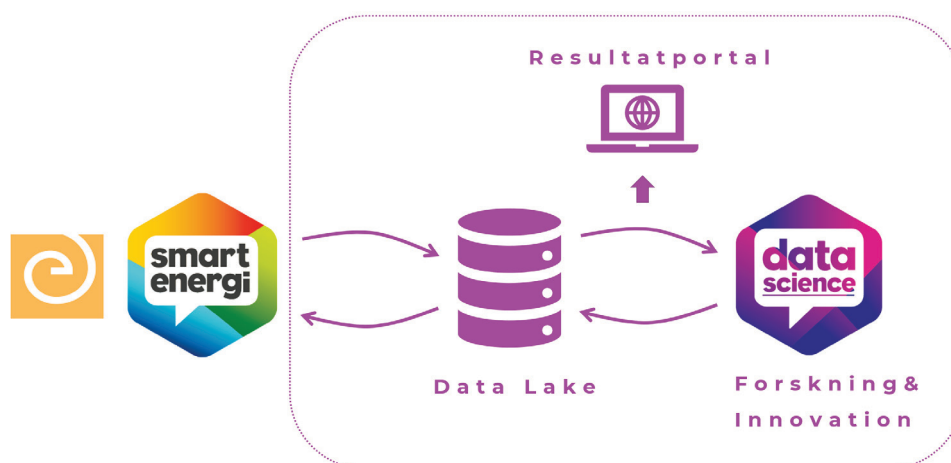


Figur 2 Projektorganisation i DS:BRAVA

2 DS:BRAVA Data Science-portal

Inom ramen för DS:BRAVA har en Data Science-portal utvecklats, för att dela dataset, resonemang, resultat, förslag till analysutmaningar, sk "spaningar".

Figur 3 nedan visar hur data från bolagen delas till en Data Lake, där analytiska resonemang och metoder tas fram. Samtliga resonemang och resultat delas i en Resultatportal, dvs. DSBRAVA Data Science Portal.



Figur 3 Data från bolagen delas till en Data Lake, där analytiska resonemang och metoder tas fram.

Materialet på Data Science-portalen består av dataset, analysutmaningar, analytiska resonemang och analytiska modeller. I dagsläget finns 37 st objekt på Data Science-portalen.

The screenshot shows the Smart Energi Data Science Portal interface. At the top, there is a search bar and a 'Filter tags' button. The left sidebar contains 'Reports' and 'Discussions BETA'. The main content area features a header with gear icons and a 'Tutorials' button. Below this is a welcome message from Filip Wastberg, dated Jan 26, 2020, with 51 views and 0 comments. The second article is 'An alternative method to identify multiple consumption "streaks" in BRAVA datalake substations' by fbenzi, dated Apr 19, 2021, with 13 views and 0 comments. The visualization shows a scatter plot of consumption data points in orange and blue, with a vertical axis labeled 'consumption'.

Figur 4 Data Science-portalen, startsida

Alla analytiska resonemang och modeller som publiceras på Data Science-portalen publiceras också med tillhörande kod i så kallad Data Science Notebooks. Det här innebär att all kod som publiceras går att användas av medlemmarna.

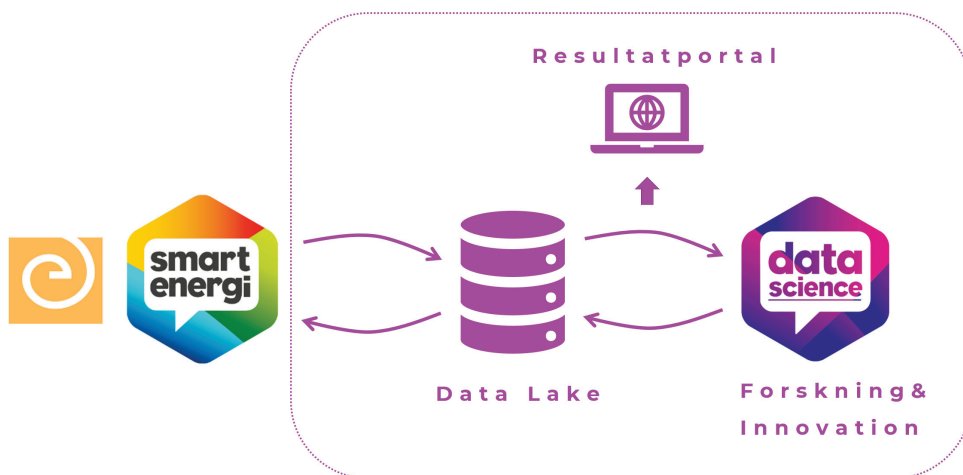
Portalen finns på följande adress:

<https://kyso.io/smartenergi/dashboard?team=smartenergi>

Information om tillgång: <https://smartenergi.org/datascienceDS:BRAVA/>

3 DS:BRAVA Data Lake

DS:BRAVA har etablerat en Data Lake för att företag som har värdefull data för forskning- och innovationssyfte ska kunna dela data med DS:BRAVA. De bolag som använder K2 ges möjlighet att per automatik och anonymt dela med sig av data som finns i K2. Data kan även läsas in med manuella överföringsmetoder, dock finns det många fördelar med att använda sig utav automatiserade datadelningsregler, både för att säkerheten per automatik följs och för att det underlättar kontinuerlig delning av data från energibolagens till innovationsprojekt. Energibolag kan själva välja vilken data de vill skicka till DS:BRAVA Data Lake.



Figur 5 Data från bolag samlas i en så kallad Data Lake

4 Avvikelseorsakstaxonomi

Det finns en mängd olika anledningar till varför en avvikelse uppstår i data som kommer från fjärrvärmekundernas värmemätare. Alla dessa orsaker måste på ett enkelt och effektivt sätt kunna kopplas till avvikelser i data, för att kunna skapa uppmärksatta dataset. Det är också viktigt att så många som möjligt använder samma etiketter för att skapa ett stort dataunderlag som kan användas för att utveckla bättre dataanalys-modeller. Därför behövs en gemensam taxonomi för orsakerna bakom avvikelserna i fjärrvärmedata - en avvikelseorsakstaxonomi. [6]

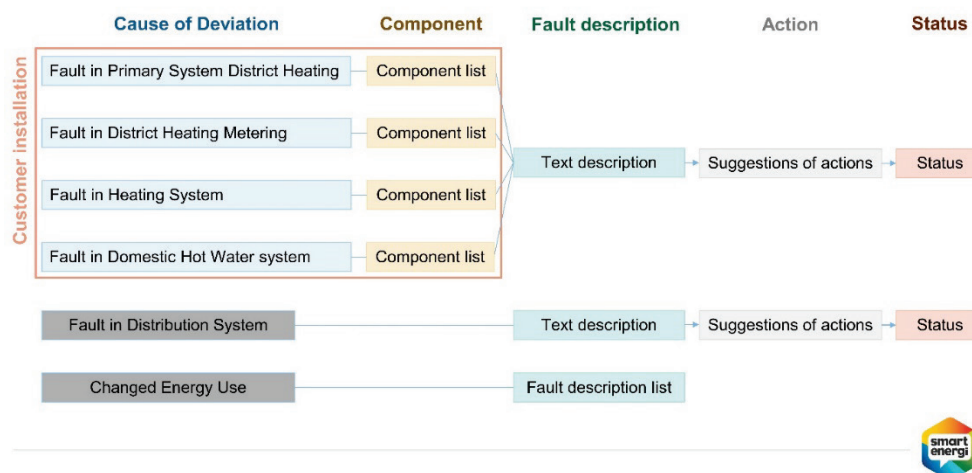
Det finns två huvudsakliga syften med en sådan taxonomi: (i) att förse fjärrvärmebranschen med ett strukturerat och standardiserat sätt att etikettera avvikelser som uppstår i fjärrvärmekunddata på grund av fel i fjärrvärmesystemen, och (ii) att hjälpa till att skapa etiketterade dataset för utveckling av prediktiva modeller som automatiskt kan upptäcka fel i fjärrvärmesystem. Taxonomin bör bestå av så många olika avvikelseorsaker som möjligt, och även innehålla information om när och var i fjärrvärmesystemet orsaken bakom avvikelsen har uppstått. Detta innebär att taxonomin måste innehålla en hel del information om de olika avvikelseorsakerna, och att den informationen måste struktureras på ett sätt som är logiskt och lätt att förstå och följa. Dessutom bör varje avvikelseorsak kunna omnämnas med en etikett bestående av ett fåtal ord och som tydligt beskriver avvikelseorsaken. Tanken är sedan att denna etikett ska användas för att etikettera dataset som innehåller avvikande datamönster, så att det är tydligt vilken orsak som ligger bakom avvikelsen i data. Med den här datan kan man sedan träna statistiska modeller att känna igen olika avvikelsemönster.

Det lättaste sättet att organisera all denna information är att arrangera den i olika nivåer. Avvikelseorsakstaxonomin som utvecklats inom detta projekt innehåller följande nivåer, som alla kommer att beskrivas i närmare detalj nedan:

- Avvikelseorsak
- Komponent
- Felorsak
- Åtgärd
- Status

I Figur 6 nedan presenteras avvikelseorsakstaxonomins övergripande struktur.

Deviation cause taxonomy



Figur 6 Avvikelseorsakstaxonominns övergripande struktur

4.1 DE FEM NIVÅERNA I AVVIKELSEORSAKSTAXONOMIN

4.1.1 Avvikelseorsak

Den första nivån i taxonomin är *Avvikelseorsak*. Denna nivå består av ett antal olika kategorier som beskriver var det problem som orsakar avvikelsen i kunddata är beläget i fjärrvärmesystemet. Det finns många olika sätt att dela in avvikelseorsakerna i kategorier. I den här taxonomin är de indelade efter funktion, alltså i vilket system eller del av ett system som orsakerna uppstår och påverkar funktionen av sagda system. Därför är avvikelseorsakerna indelade i följande kategorier i denna taxonomi:

- Fel i Primärkrets fjärrvärme
- Fel i Uppvärmningskrets
- Fel i Tappvarmvattenkrets
- Förändrad energianvändning
- Fel i Distributionsnät

De fyra första kategorierna är relaterade till fel som uppstår i kundernas byggnader, och som ger upphov till avvikelser i kunddata. Detta är också de kategorier som innehåller flest avvikelseorsaker. Den orangea rutan i figuren ovan indikerar att dessa avvikelseorsakskategorier är relaterade till fel i kundens byggnad. Kategorin *Förändrad energianvändning* är även den relaterad till kundens byggnad, i och med att det är där som en förändrad energianvändning uppstår. En förändrad energianvändning behöver dock, i många fall, inte betyda att det är något som har blivit fel i en byggnad. En förändring i energianvändningen kan exempelvis uppstå om en familj med fyra barn som spelar fotboll fem gånger i veckan och därmed duschar mycket flyttar in i en villa där det tidigare bott en äldre dam som duschar ett fåtal gånger i veckan. Detta kommer att ge en förändrad energianvändning i villan, och som troligtvis kommer att dyka upp som en avvikelse i kunddata. Detta är ju dock inte ett faktiskt *fel*, utan snarare en

förändring som är acceptabel och som kommer att bestå över tid. Därför är kategorin *Förändrad energianvändning* inte direkt relaterad till fel i kundens byggnad, som de fyra avvikelsoersakskategorierna som representerar faktiska fel i kundens byggnad är.

4.1.2 Komponent

Den andra nivån i taxonomin är *Komponent*. Denna nivå beskriver det individuella problem eller fel som orsakar en avvikelse i kunddata. De fyra kategorier som är relaterade till kundernas byggnader har på denna nivå en komponentlista som är unik för varje avvikelsoersakskategori. Kategorierna *Förändrad energianvändning* och *Fel i distributionsnät* har inga komponentlistor, utan för dessa avvikelsoersakskategorier ska användaren istället direkt fylla i en *Felbeskrivning* i den tredje nivån av taxonomin. Det finns givetvis många komponenter som skulle kunna ge upphov till fel i distributionsnätet. Det är dock troligt att många av dessa fel inte kommer att ge upphov till avvikelser i kunddata. Därför har taxonomin inte någon komponentnivå för distributionsnätet. Därför är tanken att användaren, vid de enstaka tillfällena som det uppstår fel i distributionsnätet som påverkar kunddata, ska fylla i information om felet som har uppstått i den tredje nivån av taxonomin. *Förändrad energianvändning* har ingen komponentnivå av den enkla anledningen att det inte är relevant att ha en komponentnivå för denna avvikelsoersakskategori. Om det har uppstått ett problem med en komponent i kundens byggnad som ger upphov till en avvikelse i kunddata, är tanken att detta ska täckas in av de fyra avvikelsoersakskategorierna som representerar fel i kundens byggnad.

4.1.3 Felbeskrivning

I taxonomins tredje nivå ska användaren beskriva vad som faktiskt är fel med den komponent som listas på komponentnivån (för de fyra kategorier som är relaterade till kundens byggnad). Ett exempel på detta är att en servicetekniker under ett servicebesök i en byggnad kommer fram till att avvikelsen i kunddata beror på att en reglerventil är trasig i tappvarmvattensystemet. I detta exempel är avvikelsoersakskategorin *Fel i Tappvarmvattensystemet*, Komponenterna är *Reglerventil*, och felbeskrivningen skulle vara *Trasig*. Tanken är också att användaren kommer att förse med ett antal förslag på felbeskrivningar när denna nivå fylls i, som baseras på tidigare input för denna nivå.

För avvikelsoersakskategorin *Förändrad energianvändning* finns en lista med förslag på olika felbeskrivningar. Exempel på sådana felbeskrivningar är *Förändrat uppvärmningsbehov*, *Avsiktlig avstängning*, och *Förändrad kompletterande värmeförsörjning*. För *Fel i Distributionsnät* gäller att användaren ska fylla i en (kortfattad) fritextbeskrivning om felet som har uppstått, och som har gett upphov till avvikelser i kunddata.

4.1.4 Åtgärd

Taxonomin fjärde nivå kallas för *Åtgärd*. Denna nivå representerar den handling som har utförts för att åtgärda det fel som har orsakat avvikelsen i kunddata. I och med att en förändrad energianvändning (oftast) inte är något som behöver

åtgärdas, har denna avvikelsoersakskategori inte någon åtgärdsnivå i taxonomin. Nivån består av ett antal olika alternativ av åtgärder:

- Utbytt enhet
- Reparerad
- Planerat underhåll
- Ändrad inställning
- Etc.

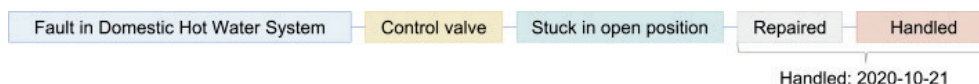
Det är också viktigt att ange när en åtgärd har blivit utförd, vilket innebär att personen som använder taxonomin måste ange ett datum och (ungefärlig) tidpunkt för när åtgärden har utförts. Detta beror på att det kan vara intressant för en dataanalytiker att jämföra data före och efter att en åtgärd har utförts, för att undersöka om mönstret i data har ändrats och huruvida avvikelsen som uppstått på grund av felet har försvunnit eller ej.

4.1.5 Status

Den femte och sista nivån i taxonomin beskriver statusen på avvikelsoersaken, det vill säga, om något har gjorts för att åtgärda problemet eller ej. I taxonomin har tre olika statusar inkluderats: *Hanterat*, *Delvis hanterat*, och *Inte hanterat*. I likhet med nivån Åtgärd är det viktigt att specificera datum och tid för statusen.

Ett exempel på att använda taxonomin

Innan de fullständiga komponentlistorna presenteras, vill vi förse läsaren med ett exempel på hur taxonomin kan användas för att registrera information om ett identifierat fel. Låt oss säga att en servicetekniker under ett servicebesök drar slutsatsen att avvikelsen i kunddata troligtvis har orsakats av en reglerventil som har gått sönder i tappvarmvattenkretsen och därmed inte fungerar längre. Serviceteknikern bestämmer sig också för att åtgärda felet direkt genom att byta ut reglerventilen, och rapporterar därefter tillbaka till fjärrvärmebolaget genom att fylla i informationen om felet i ett användargränssnitt baserat på innehållet i taxonomin. I detta exempel skulle avvikelsoersaken vara "Fault in Domestic Hot Water System", komponenten är "Control valve", felbeskrivningen är "Stuck in open position", åtgärden är "Repaired", och statusen är "Handled". Givetvis fylls även tid och datum för när åtgärden utfördes i användargränssnittet. Detta exempel är illustrerat i Figur 7 nedan.



Figur 7 Exempel på hur en sekvens av händelser kan se ut

4.2 DETALJERADE BESKRIVNINGAR AV KOMPONENTNIVÅERNA

Den andra nivån i taxonomin, Komponent, innehåller ett stort antal komponenter som kan ge upphov till avvikelser i kunddata för de fyra avvikelsoersakskategorierna som är relaterade till fel i kundens byggnad. Nedan kommer dessa fyra avvikelsoersakskategorier, samt vilka komponenter som ingår i de respektive listorna, att presenteras. De fullständiga komponentlistorna

presenteras i tabeller efter att varje individuell kategori har presenterats. I tabellerna finns även ett antal synonymer till varje komponent, som samlats in från fjärrvärmebolagen. Synonymerna står på samma rad och är separerade av "/". Det första ordet i varje rad är markerat i fetstil, vilket markerar att det är detta ord som de flesta tillfrågade fjärrvärmebolagen använder när de benämner de olika komponenterna. För varje avvikelsetorsakskategori bör det även finnas möjlighet att ange ett fritextalternativ i ett framtida användargränssnitt, om den komponent som på något sätt felat inte finns med i de olika komponentlistorna.

Fel i Primärkrets fjärrvärme

Komponenterna som tillhör Primärkrets fjärrvärme är komponenter som finns i kundens byggnad, men som inte riktigt tillhör någon av de andra avvikelsetorsakskategorierna. Det komponenterna har gemensamt är att de alla sitter på primärsidan av värmeväxlaren/na. Komponentlistan innehåller, bland annat, servisventiler som stänger av det inkommande fjärrvärmevattnet vid exempelvis underhåll eller renovering, varmhållningsventiler som möjliggör ett flöde i fjärrvärmesystemet även under varma somardagar, filter, och tömningsventiler. Se komponenter i Tabell 1 nedan.

Tabell 1 Komponenter som tillhör fel i primärkrets

Avvikelseorsakskategori	Komponent
Fel i Primärkrets fjärrvärme	Differenstryck Filter Manometerställ/Manometer/Tryckmätare/Filterkoppel Ledning/Rör Luftningsventil Servisventil Spolningsventil Strypventil/Effektbegränsningsventil Säkring elförsörjning Tömningsventil/Avtappningsventil Varmhållning servis/Rundgång

Fel i Fjärrärmätning

Värmemätaren är kopplad till fjärrvärmecentralen och mäter den mängd värme som kunden konsumerar. Mätaren mäter flödet genom fjärrvärmecentralen samt framlednings- och returtemperatur, och beräknar därefter mängden värme som kunden använder. En kommunikationsenhet kommunicerar dessa värden till fjärrvärmebolagets mätsystem, varefter de används vidare för bland annat fakturering och mätvärdesanalys. Detta innebär att det finns ett antal fel som kan uppstå längs med denna överföringskedja, och som kan uppstå hos såväl kunden som fjärrvärmebolaget. I tabellen där den fullständiga komponentlistan presenteras indikeras detta med en streckad linje – komponenter listade ovan denna linje finns i kundens byggnad, medan komponenter nedanför denna linje finns hos fjärrvärmebolaget. Se Tabell 2 nedan för komponenter.

Tabell 2 Komponenter som tillhör avvikelsekategorin Fel i fjärrvärmemätning

Avvikelseorsakskategori	Komponent
Fel i Fjärrvärmemätning	Flödesgivare Integreringsverk Kommunikationsenhet Temperaturgivare fram Temperaturgivare retur Annan störande utrustning Bristfällig kommunikation Mätvärdesinsamlingsystem Valideringssystem

Fel i Uppvärmningskrets

Uppvärmningskretsen finns inne i kundens byggnad och är det system som levererar värme till varje individuellt rum. Kretsen kan innehålla en mängd olika komponenter beroende på dess design och vilket uppvärmningsalternativ som har installerats i respektive byggnad. I taxonomin innehåller Fel i Uppvärmningskrets komponenter både på primär- och sekundärsidan av värmeväxlaren. Se komponenter i Tabell 3.

Tabell 3 Komponenter som tillhör avvikelsekategorin Fel i uppvärmningskrets

Avvikelseorsakskategori	Komponent
Fel i Uppvärmningskrets	Avstängningsventil Cirkulationspump Expansionskärl Filter Golvvärme Injusteringsventil/Strypventil Manometerställ/Manometer/Tryckmätare Påfyllningsventil Radiator Radiatorventil/Termostatventil Reglercentral/Reglering Reglerventil/Styrventil/Primärventil/Regleringsventil Shuntventil Ställmotor/Ställdon Säkerhetsventil Temperaturgivare inomhus Temperaturgivare sekundärt fram Temperaturgivare sekundärt retur Temperaturgivare utomhus Ventilationsaggregat Värmeväxlare

Fel i Tappvarmvattenkrets

Tappvarmvattenkretsen bereder och levererar varmvatten till alla tappställen i en byggnad. Tappvarmvattenkretsens design varierar beroende på byggnadens storlek och syfte. Därför finns det ett stort antal komponenter som potentiellt sett kan installeras i en tappvarmvattenkrets. Se komponenter i Tabell 4.

Tabell 4 Komponenter som tillhör avvikelsekategorin Fel i tappvarmvattenkrets

Avvikelseorsakskategori	Komponent
Fel i Tappvarmvattenkrets	Avstängningsventil Backventil Blandare med blandningsventil Blandningsventil Injusteringsventil/Strypventil Manometerställ/Manometer/Tryckmätare Reglercentral/Reglering Reglerventil/Styrventil/Primärventil/Regleringsventil Ställmotor/Ställdon Temperaturgivare blandningsventil Temperaturgivare sekundärt retur Temperaturgivare sekundärt tillopp VVC-pump/Cirkulationspump Värmeväxlare

4.3 IMPLEMENTERING AV AVVIKELSEORSAKSTAXONOMI

Målet med en implementation av avvikelseorsakstaxonomi är att kunna berika energianvändningsstatistik med information om definitiva avvikelseorsaker, insamlad på plats i fält då reparation eller service utförts. Det långsiktiga målet är här att man ska kunna utnyttja dessa avvikelseorsaker för att hitta statistiska samband och kunna utveckla bättre teoretiska modeller för att hitta problem genom energianvändningsdata.

En naturlig plattform för en implementering av avvikelseorsakstaxonomi är applikationen K2. Det som utvecklats hittills i K2 är grundläggande funktionalitet och är en grund för framtida utveckling.

Tanken är att information om avvikelseorsaker i framtiden ska samlas in när en felsökning/reparation eller service är utförd av ett energibolag. Något liknande existerar redan hos många av bolagen som lagrar information om service och utförda åtgärder, i deras faktureringsystem. Men dessa är sällan kopplade till faktiska energidata.

I K2 är alltså alla avvikelser intimt förknippade med serviceprotokoll och i K2 lagras en eller flera sådana tillsammans i protokoll som i sin tur sorteras under de anläggningar de tillhör.

Kärnan i systemet består av två delar. Den första är en specifikation av vad en avvikelseorsak är informationsmässigt. Detta omfattar definitioner för alla alternativ som är tillåtna enligt taxonomi. Dessa finns angivna i systemet som felkoder med dess tillhörande betydelser i klartext (svenska och engelska). Den andra delen är en valideringsfunktion som kontrollerar alla avvikelseorsaker som skall läggas in i systemets databas och som där säkerställer att informationen är korrekt. Detta är viktigt eftersom informationen härrör från externa källor, i slutändan från människor. Valideringen möjliggör en hög datakvalitet vilken är en förutsättning för att kunna utnyttja den på ett effektivt sätt av algoritmer.

I den implementerade funktionaliteten ingår att visualisera protokoll med tillhörande felkoder för en anläggning. I Figur 8 nedan syns ett exempel på ett sådant.

Avvikelse Id	Orsak	Komponent	Falut Beskrivning	Förändring av energianvändning	Handling	Status
0	Fel i Fjärrvärmemätning	Flödesgivare			Reparerad	Delvis åtgärdat
1	Fel i Fjärrvärmemätning	Temperaturgivare retur			Planerat underhåll	Ej åtnärmat

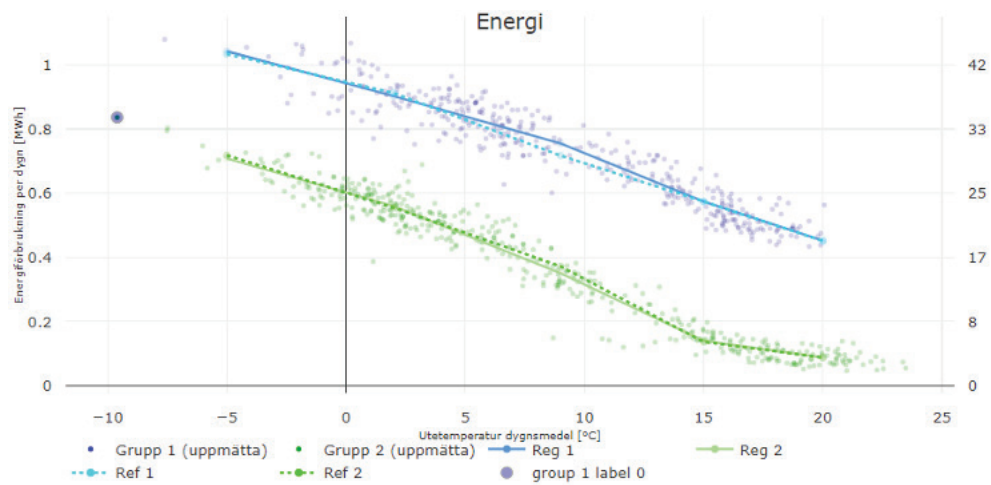
Figur 8 Exempel på protokoll och felkoder i en anläggning.

Protokoll med felkoder kan läggas in i systemet på tre olika sätt. Dessa sätt motsvarar olika nivåer av integration och ger valfrihet i hur K2 integreras med externa dataflöden.

- För enstaka protokoll finns det ett webbformulär i K2 där protokoll kan läggas till. Regler i formuläret säkerställer att informationen som läggs till följer uppsatta regler.
- Om många protokoll från många servicetillfällen läggs till kan man välja att ladda upp denna information via filuppladdning. Följande filformat stöds: EXCEL samt JSON. Systemet säkerställer även här att informationen som läggs till följer uppsatta regler.
- Automatiserad åtkomst. I framtiden kan det finnas behov av att automatisera inläggning av data. I sådana fall kan informationen läggas till via ett API som ingår i K2, skapat för detta ändamål.

För att underlätta för tredje part för att förstå hur regler samt lagring av protokoll och avvikelse-orsaker fungerar skapades ett referens Excel-ark. Arket innehåller formulär som demonstrerar ett gränssnitt för inmatning av nya avvikelseorsaker. Formuläret lägger in nya avvikelseorsaker och arket fungerar här som referens för hur informationen i avvikelseorsaker är uppbyggd. I Excel-arket går det även att läsa ut vilka inmatningsalternativ som är tillåtna, i vilka kombinationer, samt översättning till svenska och engelska för de olika felkoderna.

Syftet med integrering av felkoder i K2 är som nämnts ovan att kunna utveckla bättre teoretiska modeller. En viktig del här är att kunna markera data i de energianvändningsgrafer som ser misstänkta ut och bifoga beskrivande etiketter till dessa. Strategin är att algoritmerna kan tränas på att se statistiska genom att hitta statistiska samband mellan avvikande mätdata (med etiketter) och faktiska orsaker (avvikelsekoder). Här nedan i Figur 9 syns ett exempel på en etikett från K2. Etiketten innehåller en beskrivning samt metadata.



Figur 9 Exempel på en etikett från verktyget K2

Denna lista med etiketter visualiseras tillsammans med graferna i K2, där datapunkter med etiketter synliggörs.

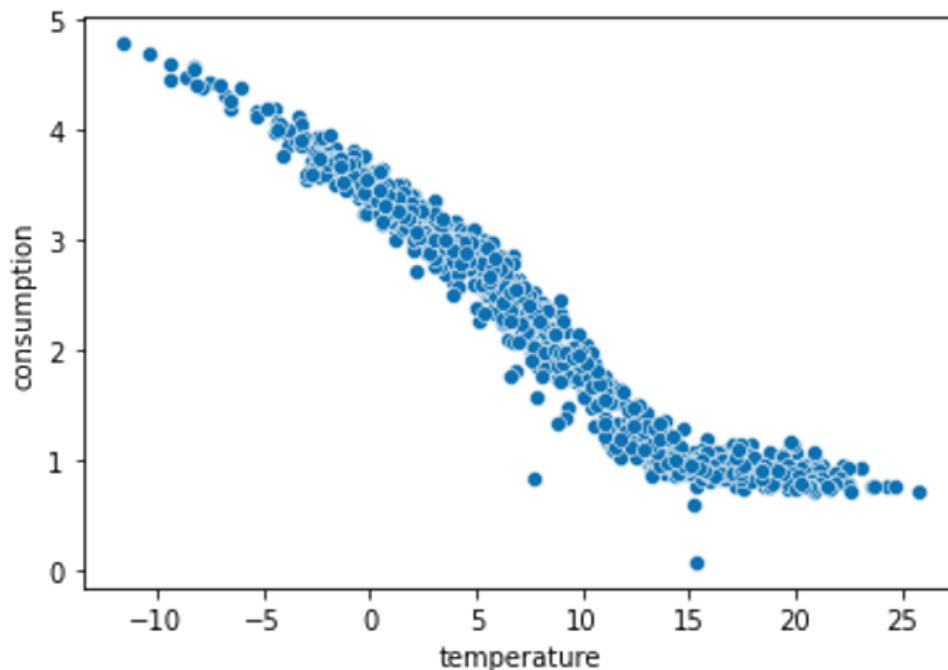
5 Utveckling av analysmodeller

Ett av huvudmålen med DS:BRAVA är att utveckla modeller för att identifiera avvikelser i fjärrvärmedata. Utgångspunkten har varit att förbättra de modeller som används för att identifiera avvikelser i K2. För att identifiera dessa avvikelser använder sig K2 av data från anläggningar/fjärrvärmecentraler, såsom flöde, flödestemperatur och returtemperatur tillsammans med externa faktorer såsom väder.

I ett klassisk maskininlärningsscenario skulle vi ha uppmärkt data med många bekräftade avvikelser, eller fel, som vi skulle kunna "träna" en algoritm på för att kunna identifiera avvikelser och fel i framtida okänd data. Inom maskininläring kallas detta för "Supervised Learning", vi *övervakar* en statistisk modell eller algoritm medan den lär sig vad som är ett "korrekt svar". Exempelvis kan vi träna en modell att känna igen foton på katter och hundar. Men vi måste då på förhand träna den matematiska modellen på av människor verifierade bilder som innehåller katter och hundar.

Den här metoden är den överlägset mest robusta och väl beprövade metoden inom maskininläring. Problemet för DS:BRAVA är att uppmärkt fjärrvärmedata, där vi vet om något har gått fel eller är en stor avvikelse, i princip inte existerar.

Metoden som K2 använder sig av är istället en typ av "Unsupervised Learning", där vi utnyttjar den starka korrelationen mellan fjärrvärmeanvändning och utomhustemperatur som ses i Figur 10.

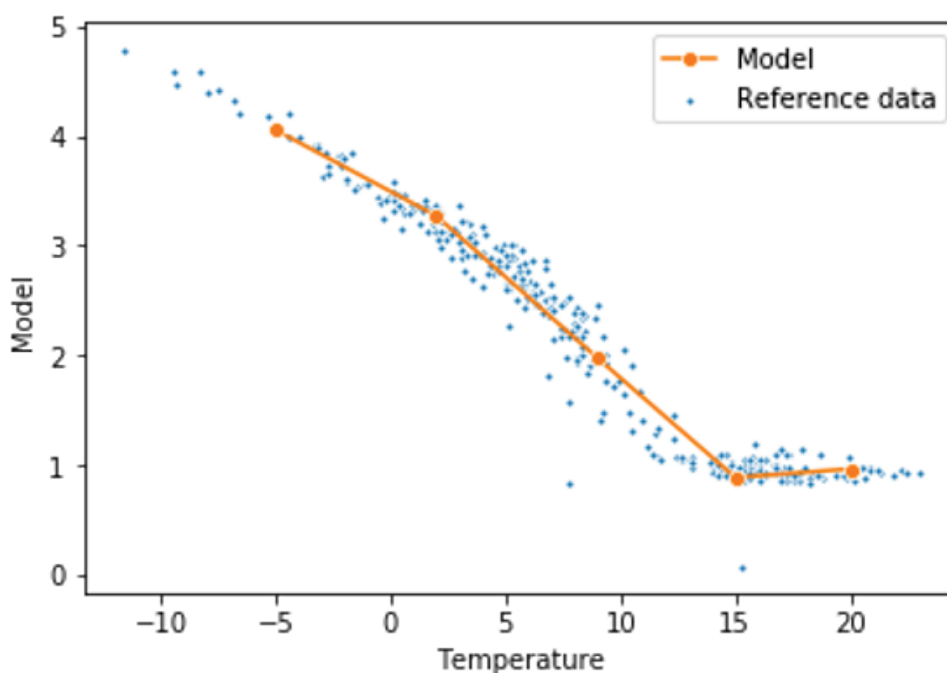


Figur 10 Korrelation mellan fjärrvärmeanvändning och utomhustemperatur

Figur 10 ovan visar korrelationen mellan fjärrvärmeanvändning och utomhustemperatur, vilket används i utformning av nya analytiska modeller och algorithmer.

Den nuvarande algoritmen i analysplattformen K2 identifierar avvikelser genom att per anläggning/fjärrvärmecentral:

1. Identifiera en referensperiod med data utan några omfattande avvikelser från den så kallade energisignaturen (relationen mellan temperatur och energianvändning). Själva identifieringen av en referensperiod görs av en analytiker och verksamhetsexpert.
2. Beräkna relationen mellan fjärrvärmeanvändning och utomhustemperatur för referensperioden med en stegvis linjär regression (piecewise linear model). Se figur 2 för exempel.
3. Använda regressionskoefficienterna från referensperioden för att utifrån temperaturerna beräkna förväntade energianvändning för perioden efter referensperiodens början.
4. Identifiera värden som befinner sig "för långt" från sina förväntade värden. "För långt" från det förväntade värdet estimeras genom en standardavvikelsemetod.



Figur 11 Referensdata och den beräknande referensmodellen

Figur 11 visar referensdata och den beräknande referensmodellen för en anläggning med en stegvis linjär regression.

Detaljer kring hur algoritmen i K2 fungerar finns på DS:BRAVAs Data Science-portal. Metoden i K2 har visat sig fungera bra eftersom den är snabb och pålitlig. Den lider dock av några problem.

1. Referensperioden väljs mer eller mindre arbiträrt och behöver mänskligt stöd för att fungera väl. För ett företag med tiotusentals anläggningar/byggnader/fjärrvärmecentraler är metoden svår att använda.
2. En stegvis linjär regression är enkel att tolka men det finns en risk att modellen inte är träffsäker på verklig data. Det kommer att bli väldigt uppenbart när energibolag vill använda modeller på timdata istället för dygnsdata.
3. Metoden att använda standardavvikelse för att identifiera avvikelser från förväntade värden är inte statistiskt robust.

I princip är modellen i K2 en prognosmodell, där man jämför prognosticerade värden mot faktiska värden för att identifiera avvikelser.

5.1 ARNOLD

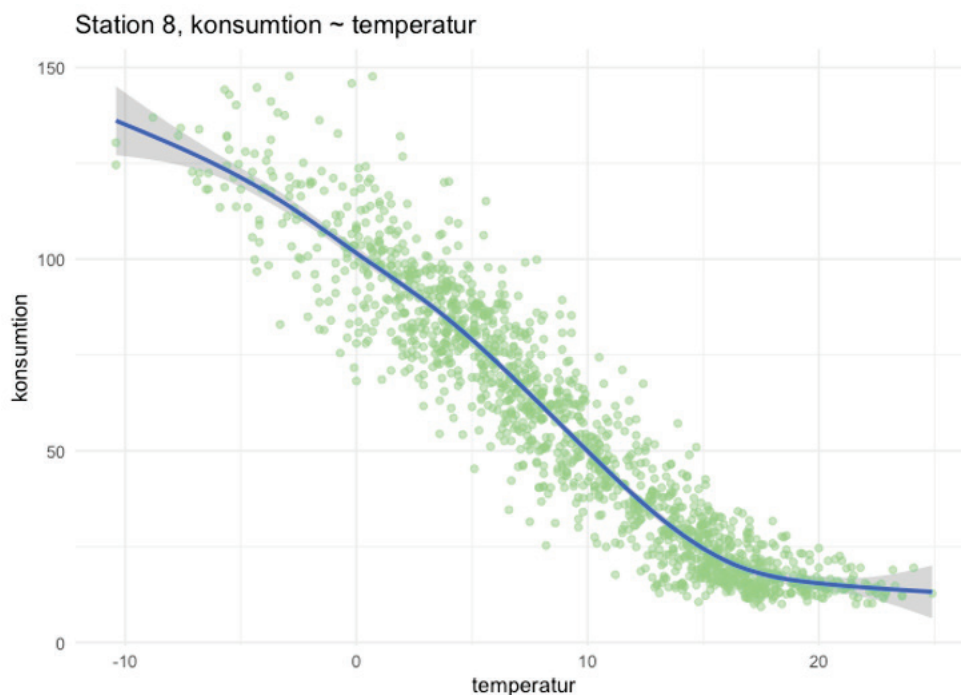
Sommaren 2020 lanserade DS:BRAVA en "analytical challenge" där medlemmar i DS:BRAVA-projektet fick möjlighet att tävla om vilken modell som kunde "slå" modellen i K2. Analysutmaningen beskrevs så här:

K2 är ett analysverktyg för att hitta avvikelser i fjärrvärmeanvändning. K2 beräknar förväntad fjärrvärmeanvändning för en anläggning/fjärrvärmecentral och jämför det sedan med faktisk användning. Om det faktiska värdet avviker för mycket från det förväntade värdet skapas ett larm. Den modell som just nu används i K2 behöver nu utvecklas för att bättre identifiera avvikelser. Om vi kan hitta en modell som prognostiserar fjärrvärmeanvändning bättre än vad K2 gör så kommer vi också kunna skapa mer träffsäkra larm. De flesta fjärrvärmebolag använder någon form av prognosmodell för budget- och produktionsplanering. Kan dessa prognosmodeller även användas för att identifiera avvikelser mer träffsäkert än nuvarande modell i K2?

Totalt lämnade fem personer in fullständiga svar på utmaningen, fem modeller lanserades och fem modeller lanserades som alternativ till K2s stegvisa regression.

1. En modell som bygger på xgboost, n populär modell för komplexa maskininlärningsproblem.
2. En tidsseriemodell
3. En segmenterad regression som är vidareutveckling av K2s stegvisa regression
4. Två bidrag använde sig av en modell kallad Generalized Additive Model (GAM) [8]. Ett av dessa bidrag vann tävlingen.

GAMs är populära för att de likt klassisk linjär regression är enkla att förstå och tolka. De har dock en fördel i att de kan användas på data där relationen mellan variabler inte är linjär, likt relationen mellan utomhustemperatur och fjärrvärmeanvändning. I Figur 12 ser ni en regressionslinje anpassad med en GAM.



Figur 12 En regressionslinje mellan konsumtion och temperatur anpassad med Generalized Additive Model

Det vinnande bidraget utvecklades till en modell som skulle kunna användas i K2 och fick namnet Arnold efter matematikern Vladimir Arnold, en rysk matematiker som varit viktig i utvecklingen av GAMs. På Data Science-portalen finns idén bakom modellen Arnold och GAMs beskriven.

5.2 REFERENSPERIOD

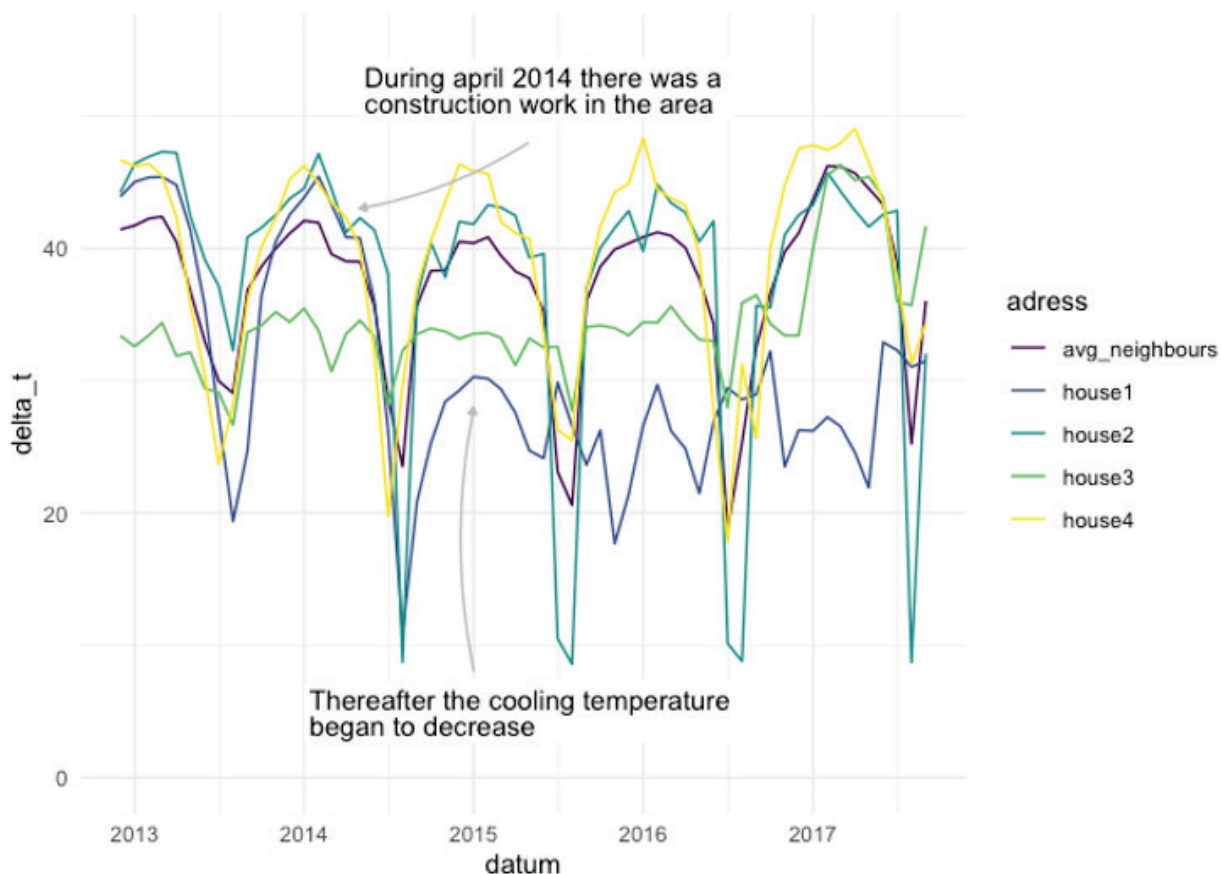
Ett annat problem för den nuvarande algoritmen i K2 är valet av referensperiod. Ett alternativ till K2s val av referensperiod är under utveckling och använder sig av en metod som heter korsvalidering. Metoden är emellertid beräkningstung och mer arbete krävs för att utvärdera den.

5.3 IDENTIFIERING AV LÄCKAGE I SERVISLEDNINGAR MED HJÄLP AV FÖRÄNDRINGSPUNKTSANALYS

Inom biomedicin är man ofta intresserad av förändringar i sekvenser av olika slag. Inom fjärrvärmeanalys är vi framför allt intresserade av förändringar i tidssekvenser, alltså när en förändring sker över tid. Inom biomedicin finns ett område som heter Förändringspunktsanalys (eng. Change point analysis) i vilket man gör just det.

Ett fjärrvärmebolag frågade DS:BRAVA om det skulle vara möjligt att identifiera läckage genom att analysera temperatur i servisledningarna från fjärrvärmesystemet in till huset. Om ett hus energianvändning, i detta fall värmeanvändning, avviker signifikant från närliggande hus bör man kunna identifiera en läcka. Fjärrvärmebolaget hade en datamängd där ett läckage hade uppstått.

I Figur 13 ser vi skillnaden mellan framledningstemperatur och returledningstemperatur, Δt , för fyra villor per månad. Under våren 2014 sjunker Δt kraftigt för ett av husen på grund av att en läcka i servisledningen uppstått.

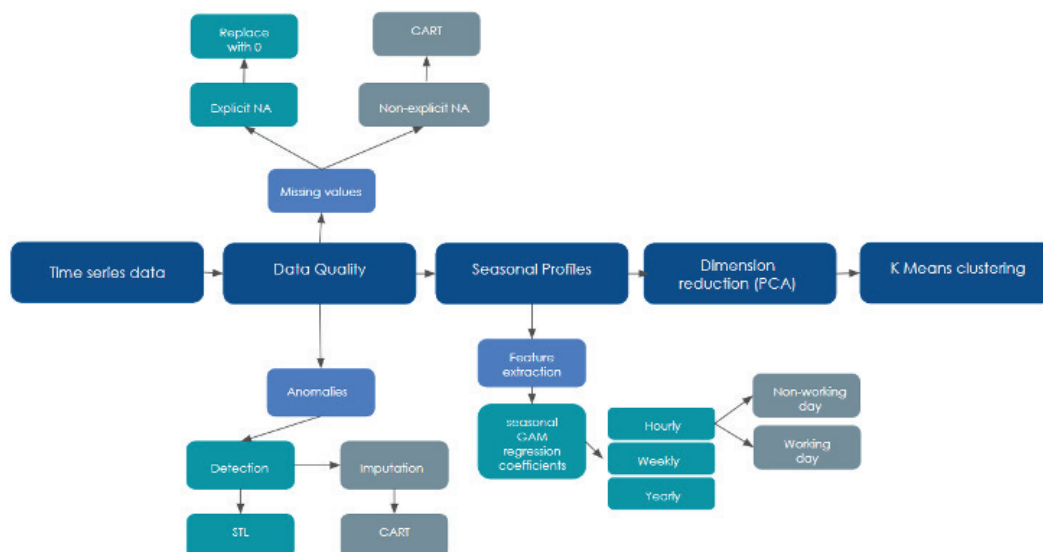


Figur 13 Skillnaden mellan framledningstemperatur och returledningstemperatur, Δt , för fyra villor per månad.

En metod kallad Bayesian Changepoint Analysis applicerades på denna datamängd och gav goda resultat. Metoden har utvärderats på några anläggningar/fjärrvärmecentraler till och fungerar bra när det är tydliga skiften i datan, men sämre om de är subtila. Även om metoden verkade lovande är den svår att utvärdera eftersom data med sanna läckor är svår att komma över.

5.4 VIRTUELLA GRANNAR

För att bättre identifiera avvikelser i data fanns det en idé inom DS:BRAVA-projektet att man skulle kunna titta på virtuella grannar för att ha referenspunkter att analysera mot. DS:BRAVA utlyste en Analysutmaning "Analytical challenge: Can you find the virtual neighbors?". Ett av svaren använde sig av tidsserieklustring enligt nedan process (Figur 14):



Figur 14 Tidsseriekustering i analysutmaningen virtuella grannar

Ett liknande system används av de modeller som Högskolan i Halmstad utvecklat.

5.5 MODELLER HÖGSKOLAN I HALMSTAD

Vi har utvecklat en data-driven metod för att möjliggöra storskalig analys av energimönster i fjärrvärmenät och hjälpa domän-experters att förstå dem bättre. Vår metod klustrar konsumentprofiler till olika grupper, extraherar representativa mönster och detekterar ovanliga konsumenter vars profil kraftigt avviker från resten av gruppen. Med vår metod, så presenterar vi den första storskaliga omfattande analysen [3] av energimönster genom en fallstudie på många byggnader i sex olika konsumentkategorier kopplade till två olika fjärrvärmenät i Sverige.

Vi har även utvecklat en avvikelsetekteringsalgoritm för övervakning av fjärrvärmesystem i stor skala. Den utgår ifrån en förmåga att identifiera en referensgrupp av fjärrvärmecentraler som är liknande en given central. Referensgruppen används med ett avvikelsetekteringsramverk som kallas "Conformal anomaly detection (CAD)" och har demonstrerats vara effektiv för att upptäcka flera olika typer av felkategorier och atypisk energianvändning.

Båda dessa arbeten finns demonstrerade i form av Jupyter Notebooks i Data science portalen.

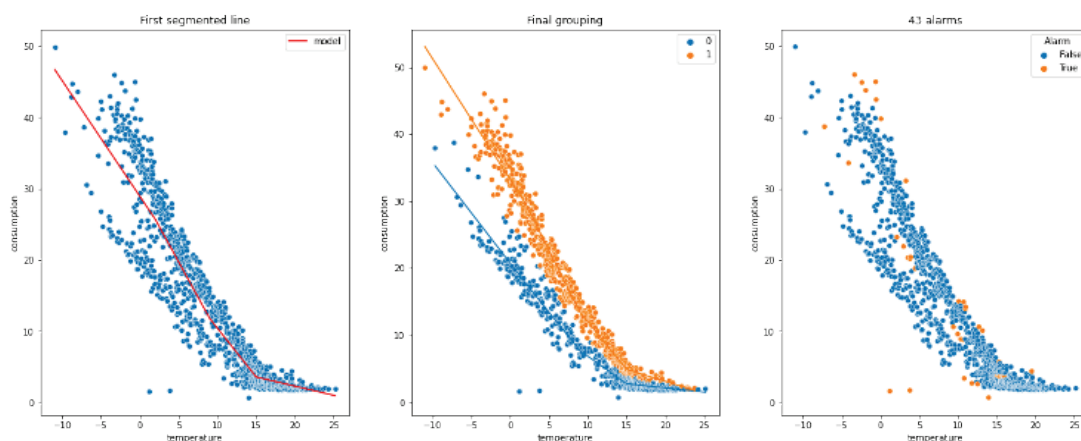
5.6 ANALYS AV ENERGIDEKLARATIONER

Energideklarationer innehåller intressant information om ett hushålls förväntade energianvändning. Inom ramen för DS:BRVA analyserades ett stort antal energideklarationer. Resultatet var intressant, men exakt hur energideklarationer ska användas inom exempelvis K2 är inte självklart.

5.7 MULTIPLA ENERGISIGNATURER

Det är inte ovanligt att fjärrvärmedata uppvisar energianvändning i flera "stråk". Ofta beror det här på olika användningsmönster under olika dagar, exempelvis på ett kontor eller en skola.

I Figur 15 finns exempel på en så kallad "tvåstråkare" och K2s identifiering av denna.



Figur 15 Exempel på identifiering av så kallade "tvåstråkare"

I dagsläget identifierar K2 stråk, som kan vara två eller flera, genom att:

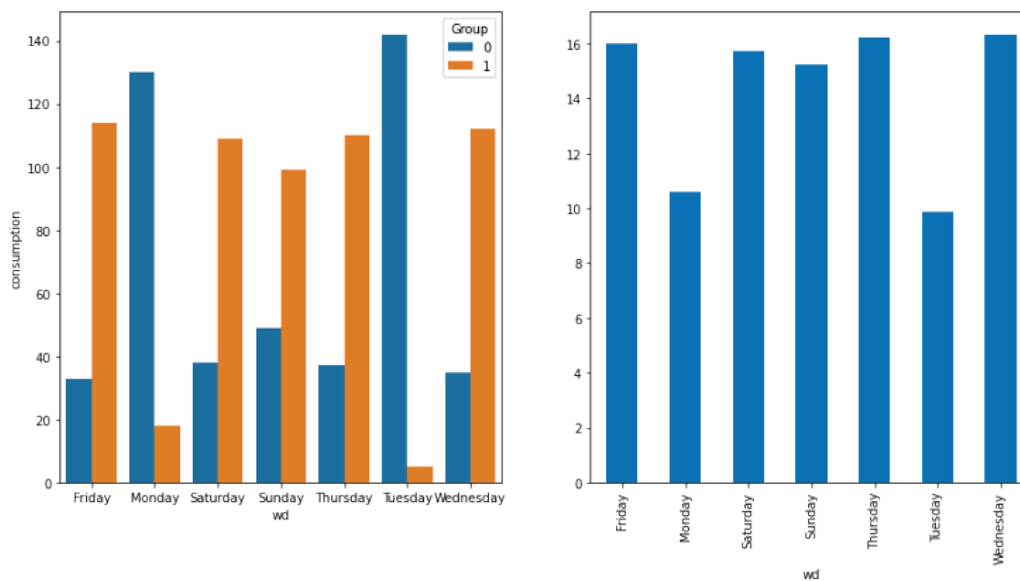
- Beräkna en stegvis linjär regression (första grafen till vänster)
- Alla punkter ovanför linjen hamnar i det övre stråket och vice versa.
- Identifiera avvikelser för respektive grupp

Metoden har fungerat väl men har några uppenbara baksidor:

Alla punkter i grupp 0 (i grafen) tillhör inte nödvändigtvis den gruppen i realiteten, utan kan lika gärna vara extremvärden från grupp 1.

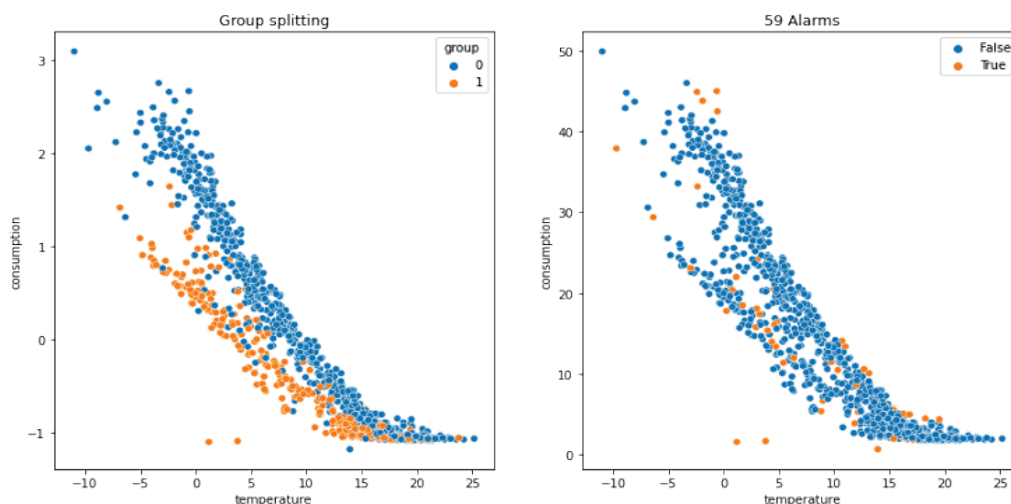
Metoden tar inte i beaktande det egentligen skälet till tvåstråkare (vilken veckodag det är).

Inom DS:BRAVA har en ny metod utvecklats för att identifiera tvåstråkare där vi använder oss av den faktiska orsaken, skillnad i veckodagar. Om vi exempelvis tittar på följande graf (Figur 16) kan vi tydligt se att fjärrvärmeanvändning skiljer sig mellan veckodagar:



Figur 16 Energiförbrukningens variation mellan olika veckodagar

Genom att gruppera observationerna på veckodag och använda oss av klustring kan vi på ett mer dynamiskt sätt hitta exempelvis tvåstråkarna:



Figur 17 Exempel på hur så kallade "Tvåstråkare" kan identifieras med hänsyn till energiförbrukningens variation

5.8 EXTREMVÄRDEN OCH IDENTIFIKATION AV ANOMALIER

Efter att ha anpassat en stegvis linjär regression på fjärrvärmedata använder K2 standardavvikelsen för att identifiera anomalier. Standardavvikelse fungerar bra på data som är normalt fördelad utan extremvärden. Men fjärrvärmedata kan ta många olika former och därför har vi inom ramen för DS:BRAVA undersökt hur vi kan använda mer robusta metoder för att identifiera extremvärden och anomalier. Två av dessa är IQR, som är en av de vanligaste metoderna för att identifiera anomalier i data, och Double MAD, som är en relativt okänd, men effektiv metod, som är robust mot olika sorters data. [4].

6 Implementering av modeller i mjukvara

6.1 RAMVERK FÖR STRINGENS I MJUKVARUTILLÄMPNING: DS:BRAVALEARN

För att möjliggöra att flertalet analytiska modeller kan tillämpas parallellt behövs ett gemensamt ramverk för att skriva kod. I DS:BRAVA har vi därför utvecklat en modul i Python som heter **DS:BRAVALearn** med vilken man kan använda modeller från ett av världens mest populära ramverk för maskininlärning, Sci-kit Learn och pandas (modul i Python) för att implementera modeller. [9] [10]

Detta ramverk kan implementeras i de analytiska applikationer, mjukvara som ska kunna ta emot nya analytiska metoder. Detta gäller t.ex. applikationen K2 som Smart Energi driver som en branschgemensam applikation, men andra applikationer på marknaden kan tillämpa samma ramverk. Samtliga energibolag som är beställare av detta projekt, DS:BRAVA, kan beställa/ implementera DS:BRAVALearn i de verktyg de använder, och därmed lägga en bra grund för att inkludera nya och parallella analytiska metoder framöver. DS:BRAVALearn utgör alltså en viktig del av en branschgemensam backend för tillämpade avancerad dataanalytiska metoder.

7 Analysvärdesmatris för avancerad analys

Värdet av att använda smarta algoritmer i fjärrvärmebranschen uppskattas av många vara högt. Göteborg Energi anger till exempel att de har efterfakturerat i snitt 1,8 Mkr/år (minst: 0,6 Mkr/år, högst: 4,4 Mkr/år) under perioden 2009-2020, mycket tack vare smart avvikelseidentifiering. Viss efter-kreditering förekommer också. Värdet av avancerad analys är dock inte begränsat till efterfakturering utan det finns stora besparingar att göra även i form av exempelvis sänkta returtemperaturer och minskat läckage.

Utvecklingen av ett verktyg vid namn Analysvärdesmatrisen har påbörjats under projektet och en preliminär version har tagits fram. Idén med Analysvärdesmatrisen är att ge en översikt över värdet av att identifiera (och åtgärda) specifika avvikelser tidigare. En god uppfattning av kostnader och potentiella besparingar kan användas till:

- Ett business-case för att ett energibolag ska börja använda avancerad analys
- Prioritera utvecklingen av algoritmer utefter vilka avvikelser som är mest värda att identifiera tidigare
- Följa upp arbetet med avvikelseidentifiering (i till exempel en dashboard)

Metodik i Analysvärdesmatrisen för att uppskatta värdet av att identifiera en avvikelse tidigare (till exempel att en flödesmätare mäter för lågt) är att först räkna ut kostnaden för ett "typiskt" fall där avvikelsen identifieras efter en viss tid, och sen räkna ut kostnaden för ett fall där avvikelsen identifieras tidigare. Skillnaden mellan dessa två fall blir värdet av att identifiera avvikelsen tidigare. En sådan metodik kräver en hel del antaganden och indata (till exempel storlek på mätarfelet, tid till upptäckt utan avancerad analys, tid till upptäckt med avancerad analys) vilket medför osäkerheter. Dessa antaganden och indata kommer dock kunna grunda sig på bättre data ju mer avancerad analys som används och ju mer data som samlas in kring olika avvikelser. Antagna indata i versionen som har tagits fram i projektet är baserade på intervjuer och uppskattningar/gissningar från representanter från energibolag, men det finns ändå en hög osäkerhet.

Den årliga besparingen beror dels på det specifika värdet av att identifiera och åtgärda varje avvikelse tidigare (kr/avvikelse), dels på frekvensen av avvikelserna (avvikelser/år). Om ett energibolag har kännedom om frekvensen för en avvikelse kan den skrivas in direkt i Analysvärdesmatrisen. I annat fall har en branschgenomsnittlig frekvens tagits fram för varje avvikelse (avvikelser/år/1000 fjärrvärmecentraler) som kan användas för att uppskatta avvikelsernas frekvens. Denna branschgenomsnittliga frekvens är i dagsläget baserad på kvalificerade gissningar från några av Energibolagens representanter, men kommer kunna bli mer noggrann ju mer avancerad analys används.

INPUT		Frekvens				
		Branchgenomsnitt	Frekvens	Specifik kostnad	Årlig kostnad	Besparing
		[#/år/1000 FV-centraler]	-Faktisk	[SEK/avvikelse]	[SEK/år]	[SEK/år]
1	INP					
2	Fjärrvärmekönet					
3	Antal fjärrvärmecentraler - näringsidkare	4000 st				
4	Antal fjärrvärmecentraler - villor	5000 st				
5	Totalt antal fjärrvärmecentraler	9000 st				
6	Total levererad värme	2000 GWh/år				
7	Uppskattad volym vatten	10000 m ³				
8	Värden för en typisk fjärrvärmecentral*					
9	Framledningstemperatur	80 °C				
10	Returtemperatur	40 °C				
11	Flöde (näringsidkare)	2,00 m ³ /h				
12	Flöde (villa)	0,20 m ³ /h				
13	Flöde (viklat medel)	1 m ³ /h				
14	Kostnadsrelaterad data					
15	Energipris (exklusive moms)*	0,65 SEK/kWh				
16	Värde i att sänka FV-temp en grad	1,5 SEK/°C/MWh				
17						
18						
19						
*Under den kalla delen på året						
					Tot [MSEK/år]	2,127
						1,293

Figur 18 Bild på analysvärdesmatris i excel

Ett exempel kan ses i Figur 18 ovan i form av en skärmdump på Analysvärdesmatrisens första blad. Verktøget är Excel-baserat och i figuren har fiktiva siffror använts för ett energibolag med 9000 fjärrvärmecentraler (4000 näringsidkare och 5000 villor), en nätvolym på 10 000 m³, framledningstemperatur på 80°C, och returtemperatur på 40°C. Den här typen av indata kan energibolagen själva fylla i. I övriga blad i Excel-filen beräknas kostnader/besparingar för de olika avvikelserna, medan sammanställningen visas på det första bladet. I det här fallet uppskattas de årliga kostnaderna för de angivna avvikelserna uppgå till drygt 2,1 MSEK/år, och besparingspotentialen av att upptäcka åtgärderna tidigare uppskattas till ca 1,3 MSEK/år. I exemplet är avvikelserna med högst besparingspotential: läckage (656 kSEK/år), flödesgivare som visar för lågt (208 kSEK/år), och styrventil som har fastnat i öppet läge (300 kSEK/år). Dessa siffror är exempel på resultat som Analysvärdesmatrisen kan ge, även om just dessa siffror har en hög osäkerhet kopplade till sig.

Som nämnts tidigare kommer Analysvärdesmatrisen kunna utvecklas mer ju mer data som samlas in. Den nuvarande versionen räknar ut kostnader och besparingar via en "bottom-up" strategi, d.v.s. genom att utgå från antaganden. I en framtid då avancerad analys används till en högre grad än idag kan besparingarna räknas ut "top-down", d.v.s. baserat på data. I ett sådant fall skulle det vara lämpligt att istället för att beskriva hur en komponent avviker ("styrventil fast i öppet läge") bara ha med komponentens namn ("styrventil") och sen räkna ut genomsnittsbesparingen för varje gång en avvikelse har identifierats och åtgärdats.

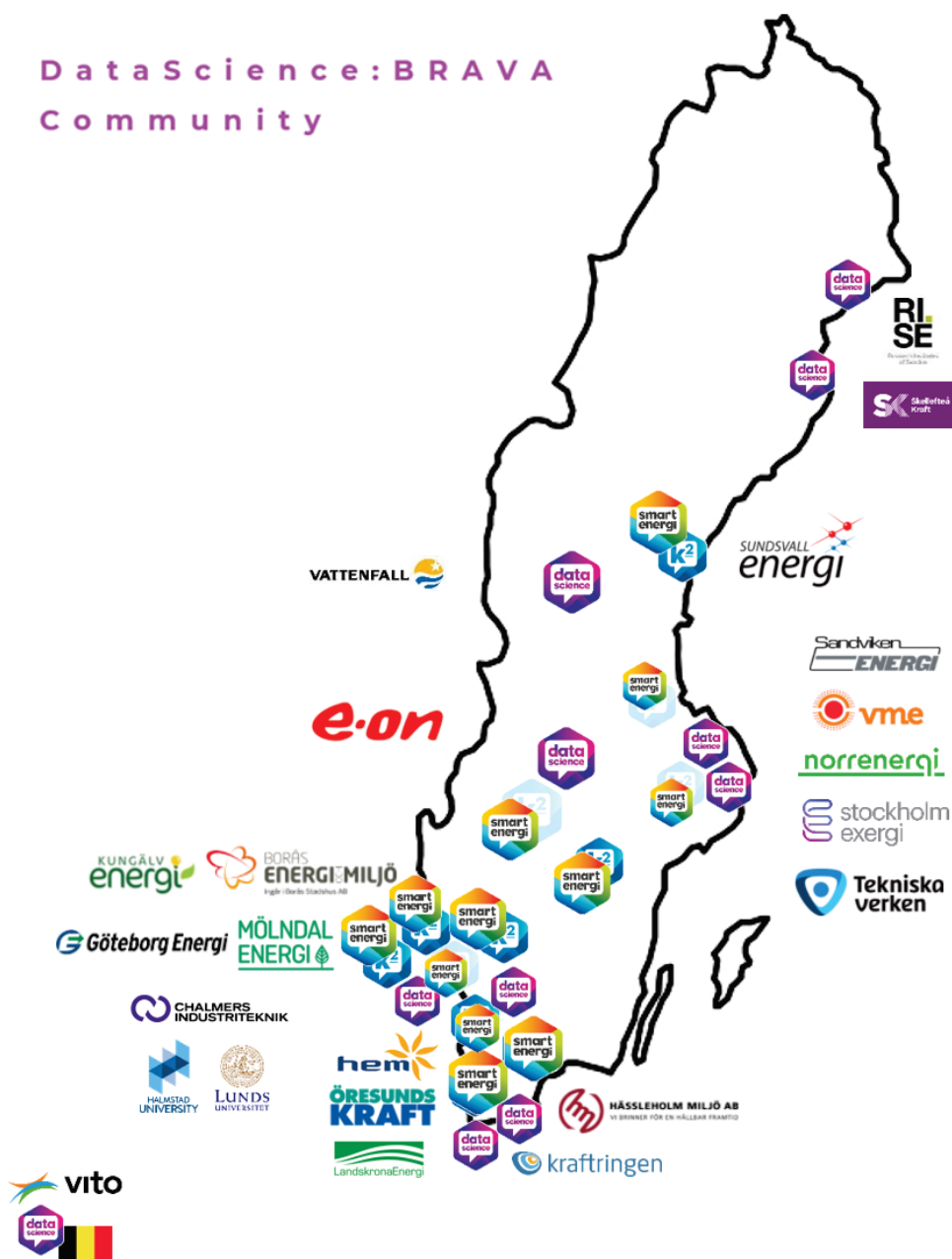
8 Community, kunskapsdelning och publika webinar/event

En viktig del av DS:BRAVA-projektet har varit att dela kunskap inom branschen. Detta har gjorts på Data Science portalen men också via öppna webinar(sk lunchsessioner), workshops i Energiforsk regi och fokusdagar anordnade av DS:BRAVA projektteam. Under dessa event har olika personer delat med sig av exempelvis en ny modell, avikelse-orsakstaxonomi med mera.

Totalt har 2 publika konferenser/fokusdagar anordnats (inkl resultatwebinaret, 11 öppna webinarer för att delge kontinuerliga projektresultat, 1 redaktionell artikel i Tidningen Energi, 2 publika workshops där Energiforsk varit sammankallande.

- DS-Community: Data Science FOCUS DAY #1, public
- DS-Community: Data Science FOCUS DAY #2, result webinar
- DS-Community: Public information + Registration instructions to Community
- DS-Community: Public Session #1 DS Portal & Data Lake Launch
- DS-Community: Public Session #2 Can you find the leak? – Yes we can!
- DS-Community: Public Session #3 Large Scale Monitoring
- DS-Community: Public Session #4 Fault Labels and why we need them
- DS-Community: Public Session #5 Data Science Lifecycle in Practice
- DS-Community: Public Session Ece Calikus Halmstad Högskola - *Cancelled*
- DS-Community: Public Session #6 Tillämpad dataanalys och applikationen K2
- DS-Community: Public Session #7 Summer Challenge - Anomaly detection using precise prognosis algorithms
- DS-Community: Public Session #8 Summer Challenge results and introducing Arnold
- DS-Community: Public Session #9 Analysvärdesmatrisen - Beräkna värdet av att arbeta med bättre metoder inom avvikelsetektering
- DS-Community: Public Session #10 Sara Månsson disputationspresentation
- DS-Community: Redaktionell artikel om samverksansprojekt inom datanaalys för energibranschen - Tidningen Energi
- DS-Community: Data set och bakgrund till case i Data Mining Kurs Högskolan i Halmstad
- DS Community: Energiforsk Workshop Vikten av branschgemensamt arbete kring Data - Datadelning, Datasäkerhet, Datastrukturer/Taxonomi/Standarder
- DS Community: Energiforsk Workshop Utvecklade kundrelationer med AI

Den Community som byggs upp, dvs de som registrerat sig som deltagare i event eller ansökt om behörighet i Data Science Portalen kommer från både energibolag och forskningsinstitut och universitet.



Figur 19 Bilden ovan beskriver Data Science Community och dess täckning.

9 Utblick för fortsatta aktiviteter

Projektet har tagit fram en infrastruktur och arbetssätt för en effektiviserad data-science-livscykel. Det är ett värde och tillgångar som skulle kastas bort om det framöver inte används och förvaltas. Utöver detta har ett flertal områden identifierats som lämpliga för fortsatt arbete.

Dessa områden **täcker ett brett** spann av möjligheter. En del områden handlar om att implementera resultat från projektet, till exempel att energibolagen börjar använda taxonomin eller att den analytiska modellen Arnold implementeras. Andra områden fokuserar på att röja undan hinder för användandet av avancerad analys, till exempel ta fram ett datadelningsramverk eller beskrivande material för datadelning och GDPR.

9.1 FÖRSLAG PÅ UPPFÖLJNINGAAKTIVITETER

I följande tabell ges en översikt av de olika aktiviteterna med förslag på ägare. En kort beskrivning av varje aktivitet följer, samt förslag på eventuell offentlig finansiering

Aktivitet	Förslag på ägare
Förvaltning av data science portal och data lake	Smart Energi
Kompetenshöjande aktiviteter i branschen	Smart Energi alt. Energiföretagen, alt. Energiforsk
Ramverk för Datadelning för forskningssyfte	Smart Energi alt. Energiföretagen, Samverksansdrivet Innovationsprojekt*
Förvaltning av taxonomin	Energiföretagen, alt. Samverksansdrivet Innovationsprojekt* för att landa ägarskapet
Förvaltning av analysvärde-matrisen	Energiföretagen, alt. Samverksansdrivet Innovationsprojekt* för att landa ägarskapet
Arbetsflöde för identifiering av avvikelser	Smart Energi alt. Energiföretagen
Tillämpning av avancerade analytiska modeller	Smart Energi och resp. Energibolag
Identifiering av läckage + testbädd	Forskningsinstitut eller universitet, Samverksansdrivet Innovationsprojekt*
Koppling mellan energibolag och fastighetsbolag	Forskningsinstitut eller universitet, Samverksansdrivet Innovationsprojekt*

*) Detta kan vara ett följdprojekt av Data Science:DS:BRAVA, t.ex. Etapp 2.

9.1.1 Förvaltning av Data Science Portal och Data Lake

Projektets delresultat och insamlade anonymiserade data finns på Data Science-portalerna och Data Laken. Hur plattformarna ska ägas, förvaltas och hur access ska hanteras är viktiga frågor framöver.

9.1.2 Kompetenshöjande aktiviteter i branschen

Projektet har byggt upp en nationellt community där många deltagit i kontinuerliga webinar och aktiviteter i syfte att dels delge branschen resultat från projektet men framför allt i syfte att höja kompetensen i branschen, vilket var ett av projektets mål. Detta community och aktiviteter bör fortsätta.

9.1.3 Datadelning för forskningssyfte

Datadelning har visat sig vara ett moment där det råder stor osäkerhet kring vad som får delas och hur det ska gå till. Ett datadelningsramverk som parter kan enas kring samt ett beskrivande material kring GDPR om vad som är tillåtet/känsligt att dela skulle kunna röja undan många hinder och snabba på utvecklingen och tillämpningen av avancerad analys.

9.1.4 Förvaltning av taxonomin

För taxonomins fortsatta användande och nyttjande är förvaltning och ägarskap viktigt. Hur skulle den kunna följas upp och eventuellt förbättras/anpassas, och vem är ansvarig för den? Hur ska spridningen av taxonomin i branschen gå till?

9.1.5 Analysvärdesmatrisen och dashboard

Hur fungerar analysvärdesmatrisen som metod för prioritering och visualisering av det sammansatta värdet av proaktivt dataanalytiska arbete? Vad behöver utvecklas och hur kan bolagen börja använda detta? En prototyp-dashboard har skapats inom projektet där avvikelser och kostnader/värden sammanställs och visualiseras. Hur fungerar den och hur ska den utvecklas?

9.1.6 Arbetsflöde för identifiering av avvikelser

En grund för att avancerad analys ska kunna användas är att det finns bra data att tillgå. Ett strukturerat arbetsflöde som resulterar i att identifierade avvikelser analyseras och etiketteras är därmed en förutsättning för ett effektivt utnyttjande av avancerad analys. Vem ansvarar för arbetssättet och ser till att det används?

9.1.7 Tillämpning av avancerad analytiska modeller

Det har utvecklats ett flertal modeller i projektet som ännu inte är implementerade. Arnold och identifiering av tvåstråkare är två exempel som nämnts tidigare i rapporten som kan implementeras framöver. Double MAD är en annan algoritm som har utvecklats men inte implementerats.

9.1.8 Identifiering av läckage + testbädd

Värdet av att identifiera läckage med hjälp av avancerad analys uppskattas vara högt. Exempelvis anger flera energibolag att man omsätter ungefär en nätvolym per år, vilket kan motsvara tusentals kubikmeter stadsvatten som behöver tillföras, behandlas och värmas upp per år. Dessutom kan läckage ge upphov till ytterligare skador och även leda till att ledningar behöver grävas upp. I avsnittet *Identifiering av läckage i servisleddningar med hjälp av Förändringspunktsanalys* ovan kan läsas om ett exempel där avancerad analys kunde användas för att identifiera en läcka som uppstår. Det skulle vara av stort värde med fler liknande fallstudier. Till exempel att testa algoritmer på fjärrvärmecentraler som ligger i områden där man har identifierat läckage med hjälp av flygtermografering. Det ska nämnas att det har funnits fall då varken algoritmer eller experter kan se att det finns en läcka bara genom att analysera data, trots att man vet att data-setet visar värden för en fjärrvärmecentral där det har funnits en läcka på servisleddningen. Ett centralt problem är att man sällan vet när en läcka uppstår. För att få bättre facitdata skulle en testbädd där man kontrollerat kan simulera drift av en fjärrvärmecentral samt läckage vara av stort värde. Brist på data och facitdata är ofta en showstopper för innovation.

9.1.9 Koppling mellan energibolag och fastighetsbolag

Vissa fastighetsbolag erbjuder sina kunder smarta tjänster, till exempel i form av översikter av energianvändning och annan data (vattenförbrukning, rumstemperatur etc) som samlas in via sensorer och som kan nås via en fastighetsportal. Energebolag erbjuder ofta sina kunder liknande tjänster, men skulle kunna engagera sig mer i hur värmen används. Här finns en möjlighet att utreda om energibolagen och fastighetsbolagen skulle kunna samarbeta för att hitta sätt att styra (eller "nudga") energianvändningen till ett smartare/effektivare användningsmönster som gynnar båda parter.

Omställningen av energisystemet mot en ökad mängd fastighetsnära energiproduktion i form av solceller, luftvärmepumpar och bergvärme leder till att konsumenter blir prosumenter med nya nettoförbrukningsmönster med nya sorters väderberoden (varmt eller kallt samt blåsigt och soligt eller stilla och molnigt väder). Framtidens energibolag behöver bli bättre på att bygga prognos och styrmodeller för att styra och **balansera** energisystemet vid olika väderförhållanden. Det finns stor potential att skapa ekonomiska värden och hållbarhetsvärden genom en bättre optimering och sam-styrning av hela energiekosystemet från produktion till prosumenter. En förutsättning för det är att energibolag och fastighetsbolag börjar dela med sig av data och information i båda riktningar på ett strukturerat sätt och att styr- och prismodeller etableras för balansering mot en effektivare och hållbarare energistyrning. En särskild möjlighet finns här att samverka mellan Allmännyttans kommunala bostadsbolag och de kommunala energibolagen då de båda är en del av våra kommuner och städer. Här finns en utmaning att designa och etablera en operativ digital samverkansmodell som baseras på data-utbyte, prognos- och styrmodeller och bransch-API som möjliggör den digitala samverkansmodellen.

9.2 POTENTIELLA OFFENTLIGA FINANSIERINGSMEDEL FÖR FORTSATTA STEG

Digitalisering och energi är två teknikområden som det satsas mycket på, och det dyker ofta upp utlysningar som kan vara intressanta för att söka finansiering till framtida aktiviteter. Nedan beskrivs förslag på offentliga program/återkommande utlysningar som skulle kunna passa aktiviteterna som nämnts hittills. Notera inget av programmen har i skrivande stund annonserat att de kommer släppa en utlysning inom kort, så det är fullt möjligt att det kommer finnas andra utlysningar som passar ännu bättre, samt att det inte kommer någon mer utlysning inom programmen.

ERA-NET SES (Smart Energy Systems). ERA-NET är ett av EU-kommissionens instrument för forskning och innovation. ERA-NET SES initiativet jobbar med fyra områden: smart power grids, integrated regional energy systems, flexible heating and cooling systems, och smart services. Finansiering går till transnationella projekt. I den senaste ERA-NET SES utlysningen fanns krav på minst tre oberoende parter från minst två olika länder. Exempel på aktiviteter skulle kunna vara att sprida taxonomin och arbetsflöde för identifiering av avvikelser utanför Sverige.

TERMO. Energimyndighetens forsknings- och innovationsprogram TERMO har finansierat DS:BRAVA-projektet med 54% av det totala beloppet. Om fler utlysningar släpps inom programmet eller om det blir en uppföljning skulle det vara ett naturligt ställe att söka ifrån för fortsättningsaktiviteter.

AI i klimatets tjänst. Detta är en utlysning från Vinnova, Formas, Energimyndigheten och Rymdstyrelsen med fokus på minskning av växthusgaser som har släppts i två omgångar. Denna utlysning passar alltså bara projekt som kan förväntas ge upphov till minskningar av växthusgaser. Eftersom fjärrvärmebranschen har som mål att vara fossilfri 2030 kan det vara svårt att argumentera för att en digitalisering ger upphov till stora utsläppsminskningar. Däremot är det inte omöjligt att det i framtiden blir ökat fokus på att även vara resurseffektiv med användandet av biomassa som bränsle, och då kan digitalisering av fjärrvärmesektorn (och därmed en förväntad minskning av returtemperaturen och ett effektivare system) komma att bli aktuellt för den här eller liknande utlysningar.

Energiforsk / Futureheat. Under förutsättning att Energiforsk kommer med fortsatta utlysningar.

10 Referenslista

- [1] Svt nyheter, " Millimetersmå läckor under jord upptäcks genom flygspaning" 10 februari 2020. [Online]. Available: <https://www.svt.se/nyheter/lokalt/stockholm/millimetersma-lackor-under-jord-hittas-med-hjalp-av-flygspaning>.
- [2] E. Calikus, S. Nowaczyk, A. Sant'Anna, H. Gadd, S. Werner, "A data-driven approach for discovering heat load patterns in district heating," *Applied Energy* 252, vol. 113409, 2019.
- [3] S. Farouq, S. Byttner, MR. Bouguelia, N. Nord, H. Gadd, "Large-scale monitoring of operationally diverse district heating substations: A reference-group based approach," *Engineering applications of artificial intelligence* , nr 90, 2020.
- [4] A. Akinshin, "DoubleMAD outlier detector based on the Harrell-Davis quantile estimator," 22 juni 2020. [Online]. Available: <https://aakinshin.net/posts/harrell-davis-double-mad-outlier-detector/>
- [5] H. Gadd, S. Werner, "Fault detection in district heating substations," *Appl. Energy*, vol. 157 (2015), pp. 51-59. <https://doi.org/10.1016/j.apenergy.2015.07.061>
- [6] S. Månsson. et al., "A taxonomy for labeling deviations in district heating customer data," *Smart Energy*, Volume 2 (2021), 100020, ISSN 2666-9552. <https://doi.org/10.1016/j.segy.2021.100020>
- [7] S Månsson, M. Thern, P-O. Johansson Kallioniemi, K. Sernhed, "A Fault Handling Process for Faults in District Heating Customer Installations," *Energies* vol. 14, no. 11 (2021). <https://doi.org/10.3390/en14113169>
- [8] S. N. Wood, *Generalized Additive Models: An Introduction with R*, Second Edition, Chapman and Hall/CRC, 2017.
- [9] W. McKinney et al., "Data structures for statistical computing in python," *In Proceedings of the 9th Python in Science Conference*, p. 51-56, 2010.
- [10] H. Wickham. et al., "Welcome to the tidyverse," *Journal of Open Source Software*, 4(43), 2019, 1686 <https://doi.org/10.21105/joss.01686>

BRANSCHSAMARBETE FÖR AVANCE- RAD ANALYS AV VÄRMEDISTRIBU- TION OCH UPPVÄRMNINGSBEHOV

Målet med det här projektet är att öka den analytiska förmågan i fjärrvärmebranschen och hur man kan etablera en effektivare och snabbare cykel för att arbeta fram nya och förbättrade metoder inom avancerad analys- och maskinlärning/AI. Genom bättre metoder blir det lättare att hitta avvikelser och förbättringsmöjligheter inom hela värmedistributionen.

Inom ramen för projektet har en taxonomi för etikettering av data tagits fram. Taxonomin skapar stringens för branschen och gör det enklare för företag att märka upp fel-åtgärder-orsaker i fjärrvärmedata för att kunna träna maskinlärningsmodeller. Fjärrvärmedata har samlats från energiföretag, framförallt från analysapplikationen K2 till en gemensam dataplattform kopplad till en data science-portal. Projektet har också validerat analysmodellerna i den branschgemensamma applikationen.

Ett nytt steg i energiforskningen

Energiforsk är en forsknings- och kunskapsorganisation som samlar stora delar av svensk forskning och utveckling om energi. Målet är att öka effektivitet och nyttiggörande av resultat inför framtida utmaningar inom energiområdet. Vi verkar inom ett antal forskningsområden, och tar fram kunskap om resurseffektiv energi i ett helhetsperspektiv – från källan, via omvandling och överföring till användning av energin.