

Energimyndighetens titel på projektet – svenska Branschsamarbete för avancerad analys för framtidens uppvärmning (DS:BRAVA)	
Energimyndighetens titel på projektet – engelska	
Universitet/högskola/företag Solita/Halmstad Högskola/Chalmers Industri teknik/Smart Energi	Avdelning/institution
Adress	
Namn på projektledare Maria Hansson	
Namn på ev övriga projektdeltagare Filip Wästberg, Ida Lundholm Benzi, Federico Benzi, Anders Tholén, Simon Löfwander, Kristoffer Hellqvist, Per Grosskopf, Sara Månsson, Rikard Edland, Joakim Wigström, Stefan Byttner, Ece Calikus, Shiraz Farouq	
Nyckelord: 5-7 st AI, Data Science, Avancerad Analys, Felkoder, Taxonomi, Standardisering, Fjärrvärme	

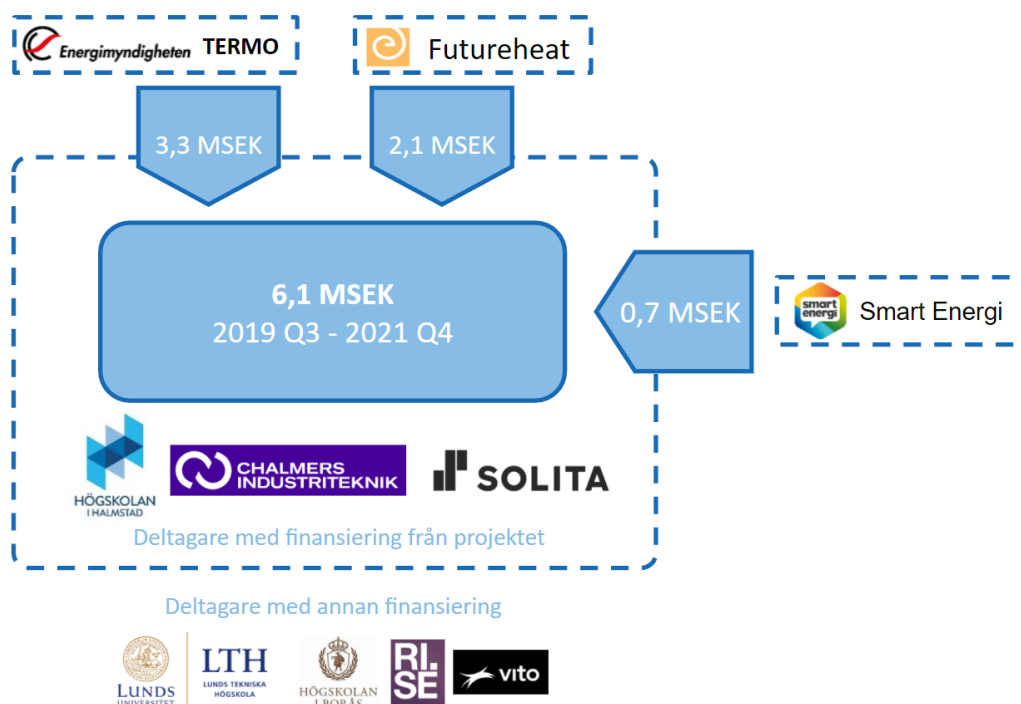
Förord

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 mjukvarututveckling.

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 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.



Personal från energibolag i
Sverige i form av Data Science
Community deltagare

Innehållsförteckning

Sammanfattning	4
Summary	5
Inledning/Bakgrund	6
Avsaknad av data med hög kvalitet	6
Avsaknad av kompetens och verktyg	7
Mål och syfte	7
Genomförande	7
Data Science råd	7
Avvikelse-orsakstaxonomi	8
Data Science-portalen	8
DS:BRAVA Data Lake	8
Utveckling av analysmodeller	8
Implementering av modeller	8
Kunskapsdelning och publika webinar/event	9
Analysvärdesmatris för att identifiera värdet av avancerad analys	9
Resultat	9
Avvikelse-orsakstaxonomi	9
Implementering av Avvikelse-orsakstaxonomi	10
Data Science-portalen	10
DS:BRAVA Data Lake	11
Implementering av analysmodeller	11
Kunskapsdelning och publika webinar/event	11
Analysvärdesmatris	11
Diskussion	12
Publikationslista	13
Referenser, källor	13
Bilagor	14

Sammanfattning

Projektet branschsamarbete för avancerad analys för framtidens uppvärmning (också kallat för DS:BRAVA - Data Science: Branschgemensam analys) hade som syfte att ta fram algoritmer och datastrukturer för att rekommendera energieffektiviserande åtgärder och åtgärda fel hos distributörer och prosumenter för att uppnå högre effektivitet i energisystemet. Arbetet inom DS:BRAVA har framför allt fokuserat på att göra det enklare för fjärrvärmebolag att hitta fel i fjärrvärmesystemet.

Projektet har i linje med målen först och främst fokuserat på att ta fram analysmodeller och att göra data tillgänglig för samarbetsparterna. Parallellt med det här har projektet identifierat en i princip avsaknad av uppmärkt träningsdata, som är avgörande för framtagande av analysmodeller. För att tackla problemet har DS:BRAVA bland annat tagit fram en branschgemensam Avvikelse-orsakstaxonomi för uppmärkning av träningsdata. Det har även tagits fram en analysvärdesmatrix för att påvisa det faktiska värdet av avancerad analys.

DS:BRAVA har etablerat en Data Science-portal där analysmodeller, analysutmaningar, dataset och ett 30-tal så kallade computational notebooks publicerats. Till portalen har också en datalake skapats dit data från energibolagen kan skickas för analys inom DS:BRAVA. Under projektet har också ett antal lunchseminarier anordnats mellan energibolag och universitet. Dessa har varit ett uppskattat kompetensutbyte mellan olika aktörer i fjärrvärmebranschen.

Projektet har tagit fram ett datadelningsramverk som bolagen har kunnat använda som stöd i steget att dela data till ändamål inom forskning och innovation.

För att öka kunskapsspridningen inom branschen har DS:BRAVA anordnat flertalet webinarier och evenemang där olika delar av arbetet presenterats och diskuterats.

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.

Inledning/Bakgrund

Enligt rapporten *Fault detection in district heating substations*, H. Gadd, S. Werner, 2015 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.

För att forcera utvecklingen av avvikelседetektering med hjälp av maskininläring och artificiell intelligens, men också för att låta andra energibolag (som inte är med i Smart Energi) ta del av kompetensutvecklingen, lanserades projektet Branschsamarbete för avancerad analys för framtidens uppvärmning (BRAVA).

Avsaknad av data med hög kvalitet

De statistiska modeller som brett kallas för AI, som ligger bakom exempelvis självkörande bilar och ansiktsgenkänning är i regel utvecklade på data med tydliga etiketter (eng. labels). 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 maskininläring.

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 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.

Avsaknad av kompetens och verktyg

Men det går att komma framåt även utan etiketter. Problemet är ofta snarare kunskap och kompetens. Att främja samarbete mellan universitet, företag och industrispecialister är därför av stor vikt. DS:BRAVA-projektet har därför haft som fokus 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 integrering via samarbetet Smart Energi, som erbjuder mjukvaran K2. K2 gör det möjligt för företag att snabbare och mer effektivt kunna hitta avvikelser.

Det här resulterar i en symbios där intäkter till Smart Energi och K2 finansierar och stimulerar samarbete mellan olika aktörer inom fjärrvärmebranschen; dessa kan sedan ta nytta av samarbetet i mjukvaran K2.

Mål och syfte

Målet för studien var bland annat att etablera en data science plattform där dataset, analysutmaningar och analysmodeller kunde publiceras. Syftet med dessa analysmodeller var att kunna erbjuda färdig mjukvara, i form av så kallade computational notebooks, för att kunna nå högre energieffektivitet för energidistributörer och prosumenter.

Utifrån energibolagens prioritering var målet att 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 avancerade 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.

Genomförande

Data Science råd

För att överse och prioritera arbetet inom DS:BRAVA etablerades vid projektets uppstart ett Data Science Råd med representanter från både Energibolag, forskning/akademi och data science/mjukvaruexpertis. Syftet med rådet var att diskutera och prioritera arbetet inom DS:BRAVA.

Avvikelse-orsakstaxonomi

BRAVA inleddes med workshops mellan samarbetets olika aktörer. Syftet var att kartlägga var hos energibolag det fanns mest resurser att spara genom användandet av avancerad dataanalys. Läckage i fjärrvärmenätet ansågs av många vara de mest kostsamma felen som kan uppstå. Ett läckage kostar energibolaget resurser och riskerar att skapa en dålig relation till kunden om det inte upptäcks i tid. Under dessa workshops var det tydligt att många företag saknade en uppfattning om hur många fel de hade i sitt fjärrvärmenät, läckage inkluderat. De saknade också en metod för att märka upp fel på ett systematiskt sätt.

Efter diskussioner med BRAVAs referensgrupp inleddes ett arbete med BRAVAs Avvikelse-orsakstaxonomi. Under ledning av Sara Månsson, doktorand från Lunds universitet, gjordes systematiska studier av vad som kan orsaka avvikelser i kunddata och hur dessa kan kategoriseras.

Data Science-portalen

Men det går att komma vidare utan uppmärkt data, framför allt med hjälp av så kallad Unsupervised Learning, som inte kräver uppmärkt data. För att enkelt kunna samarbeta mellan experter, fjärrvärmebolag och universitet skapades en så kallad Data Science-portal. På portalen kan samarbetsparter publicera och hämta dataset, analysutmaningar, analytiska resonemang och analytiska modeller.

DS:BRAVA Data Lake

För att göra det möjligt för företag att dela med sig av data enkelt och anonymt skapades en så kallad Data Lake dit fjärrvärmeföretag kan publicera data för forskningsändamål.

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 mätstationer, såsom flöde, framledningstemperatur och returtemperatur tillsammans med externa faktorer såsom väder.

På Data Science-portalen började deltagare i DS:BRAVA utveckla nya modeller som på ett mer träffsäkert sätt kan identifiera avvikelser. Utöver avvikelседetektering började medverkande ta fram metoder för att identifiera läckage (där uppmärkt data finns tillgänglig), algoritmer för virtuella grannar och analys av annan data relevant för branchen.

Implementering av modeller

För att enklare kunna implementera de modeller som utvecklas på Data Science-portalen började DS:BRAVA arbeta med utveckling av ett open source ramverk för implementering av avancerade analysmodeller i fjärrvärmebranschen. Arbetsnamnet för implementeringsramverket var BRAVALearn.

Kunskapsdelning och publika webinar/event

En viktig del av BRAVA-projektet var att dela kunskap inom branschen. För att göra det här började BRAVA anordna öppna webinarier (sk lunch-sessioner), workshops i Energiforsk regi och fokusdagar anordnade av DS:BRAVA projektteam.

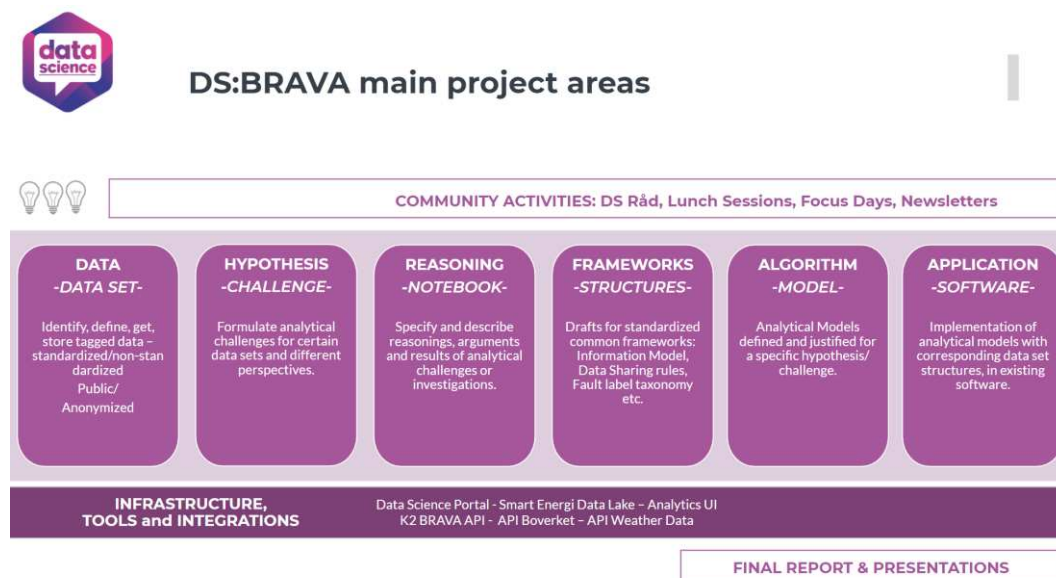
Analysvärdesmatris för att identifiera värdet av avancerad analys

För att få företag att investera i avancerad analys behöver företag kunna kvantifiera värdet av analysen. För att ge företag verktyg att göra just detta började DS:BRAVA ta fram en analysvärdesmatris för företag att använda.

En god uppfattning av värdet av att identifiera avvikelser - i form av till exempel en analysvärdesmatris - 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)

Följande bild beskriver projektets huvudsakliga aktivitetsområden:



Resultat

Avvikelse-orsakstaxonomi

Den slutgiltiga avvikelse-orsakstaxonomi är uppdelad i olika nivåer där avvikelser och fel kan uppstå och ska underlätta för tekniker och analytiker att kunna märka upp fel och

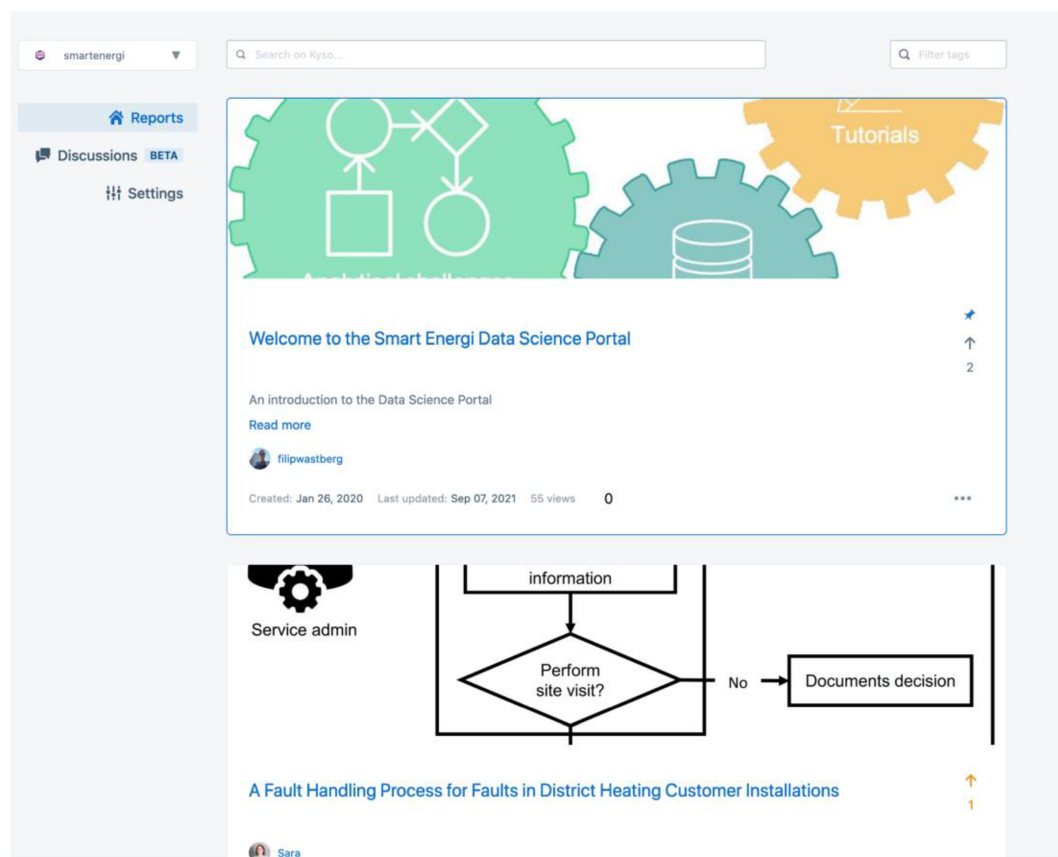
avvikelser i fjärrvärmedata. Arbetets resultat var en stor del av Sara Månssons avhandling¹ och har resulterat i vetenskapliga artiklar^{2 3}.

Implementering av Avvikelse-orsakstaxonomi

För att kunna använda taxonomin rent praktiskt implementerades den i mjukvaran K2 och finns nu tillgänglig för K2s användare. Det här gör att analytiker själva kan märka upp data med avvikelser och på så sätt skapa träningsdata för framtida analytiska modeller.

Data Science-portalen

Data Science-portalen började fort fyllas med intressanta dataset, utmaningar och analytiska modeller. I dag finns det 37 st objekt på Data Science-portalen som ser ut så här:



¹ Spot the difference! On the way towards automated fault handling in district heating buildings, Månsson, S

² A taxonomy for labeling deviations in district heating customer data
Månsson, S., Lundholm Benzi, I., Thern, M., Salenbien, R., Sernhed, K. & Johansson Kallioniemi, P-O., Science Direct Smart Energy. Volume 2, May 2021 2021

³ Faults in district heating customer installations and ways to approach them: Experiences from Swedish utilities, Sara Månsson, Per-Olof Johansson Kallioniemi, Marcus Thern, Tijs Van Oevelen, Kerstin Sernhed, Science Direct Energy Volume 180, 1 August 2019

Portalen finns på följande adress (kräver inloggning):

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

DS:BRAVA Data Lake

Under projektet har BRAVA byggt en data lake till vilken företag som använder K2 kan bidra med anonym data för forskningssyfte. Databasen innehåller i dagsläget miljontals rader data och har varit viktig i arbetet med att utveckla nya algoritmer och analytiska modeller.

Implementering av analysmodeller

Flera modeller har tagits fram på Data Science-portalen. För att underlätta implementering av dessa modeller har ett ramverk, BRAVALearn, tagits fram. Det har också gjorts att modeller kunnat implementeras i mjukvaran K2.

Flera av modellerna som tagits fram har också blivit vetenskapliga artiklar i huvudsak från Högskolan i Halmstad.⁴

Kunskapsdelning och publika webinar/event

Totalt har ett tiotal publika webinar genomförts som antingen lunch-sessioner eller öppna evenemang. Därtill har BRAVA bidragit med en artikel till Tidningen Energi samt genomfört en workshop med Energiforsk om Vikten av branschgemensamt arbete kring Data.

Analysvärdesmatris

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.

Analysvärdesmatrisen kommer 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 och kvalificerade gissningar (baserat på diskussioner med experter från energibolagen). I en framtid då avancerad analys används till en högre grad än idag kan besparingarna räknas ut ”top-down”, d.v.s. baserat på statistik över faktiska besparingar som har samlats in genom ett systematiskt användande av exempelvis avvikelse-orsakstaxonomin. Läckage och flödesmätarfel är exempel på avvikelser som den nuvarande versionen av analysvärdesmatrisen uppskattar som extra värdefulla att upptäcka tidigt.

⁴ Large-scale monitoring of operationally diverse district heating substations: A reference-group based approach. S Farouq, S Byttner, MR Bouguelia, N Nord, H Gadd. Engineering applications of artificial intelligence 90. 2020.

Diskussion

Då detta projekt syftade till att ta fram en infrastruktur och ett arbetssätt för en effektiviserad data-science-livscykel, dvs. hur innovation kring Data Science och AI kan tillämpas i energibolagens verktyg processer och rutiner är det viktigt att resultaten övergår till vidare implementering och användning. Samtliga resultat finns beskrivna på den Data Science Portal projektet etablerat och som finns tillgänglig när projektet avslutats. Utöver faktiska resultat, redo att implementeras, har ett flertal områden identifierats som lämpliga för fortsatt arbete.

Förslagen för fortsatt arbete 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.

Förslag på uppföljningsaktiviteter

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 analysvärdematrisen	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.

Publikationslista

Faults in district heating customer installations and ways to approach them:
Experiences from Swedish utilities, Sara Månsson, Per-Olof Johansson
Kallioniemi, Marcus Thern, Tijs Van Oevelen, Kerstin Sernhed, ScienceDirect
Energy Volume 180, 1 August 2019

A taxonomy for labeling deviations in district heating customer data, Månsson, S.,
Lundholm Benzi, I., Thern, M., Salenbien, R., Sernhed, K. & Johansson
Kallioniemi, ScienceDirect Smart Energy, Volume 2, May 2021

Large-scale monitoring of operationally diverse district heating substations: A
reference-group based approach. S Farouq, S Byttner, MR Bouguelia, N Nord, H
Gadd. Engineering applications of artificial intelligence 90. 2020.

Mondrian conformal anomaly detection for fault sequence identification in
heterogeneous fleets. Shiraz Farouq, Stefan Byttner, Mohamed-Rafik Bouguelia,
Henrik Gadd. Neurocomputing, Volume 462, 2021, Pages 591-606, ISSN 0925-
2312.

Licentiatavhandling:

Farouq, S. (2019). Towards large-scale monitoring of operationally diverse
thermal energy systems with data-driven techniques, Licentiate dissertation,
Halmstad University Press.

Referenser, källor

H. Gadd, S. Werner, Fault detection in district heating substations, Appl.
Energy, 157 (2015), pp. 51-59, 10.1016/j.apenergy.2015.07.061

A data-driven approach for discovering heat load patterns in district heating. E Calikus, S Nowaczyk, A Sant'Anna, H Gadd, S Werner. *Applied Energy* 252, 113409. 2019.

<https://www.svt.se/nyheter/lokalt/stockholm/millimetersma-lackor-under-jord-hittas-med-hjalp-av-flygspaning>

Large-scale monitoring of operationally diverse district heating substations: A reference-group based approach. S Farouq, S Byttner, MR Bouguelia, N Nord, H Gadd. *Engineering applications of artificial intelligence* 90. 2020.

<https://aakinshin.net/posts/harrell-davis-double-mad-outlier-detector/>

Wood S (2017). *Generalized Additive Models: An Introduction with R*, 2 edition. Chapman and Hall/CRC.

McKinney, W. & others, 2010. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*. pp. 51–56.

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). “Welcome to the tidyverse.” *Journal of Open Source Software*, 4(43), 1686. doi: 10.21105/joss.01686.

Faults in district heating customer installations and ways to approach them: Experiences from Swedish utilities, Sara Månsson, Per-Olof Johansson Kallioniemi, Marcus Thern, Tijs Van Oevelen, Kerstin Sernhed, *ScienceDirect Energy* Volume 180, 1 August 2019

A taxonomy for labeling deviations in district heating customer data, Månsson, S., Lundholm Benzi, I., Thern, M., Salenbien, R., Sernhed, K. & Johansson Kallioniemi, *ScienceDirect Smart Energy*, Volume 2, May 2021

Bilagor

1. Administrativ bilaga