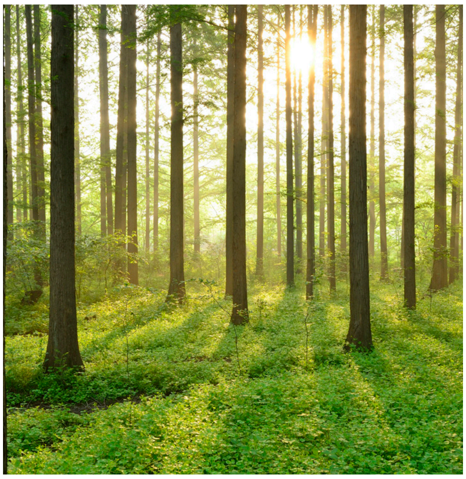
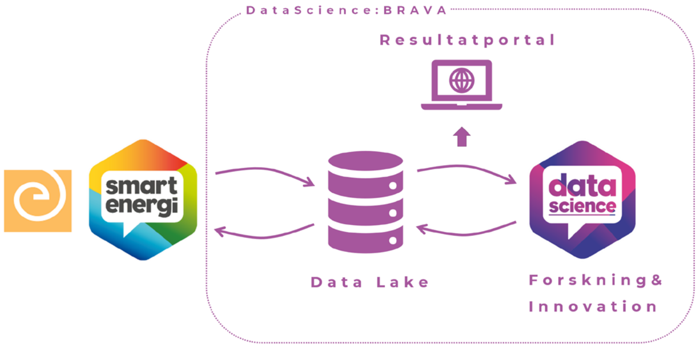
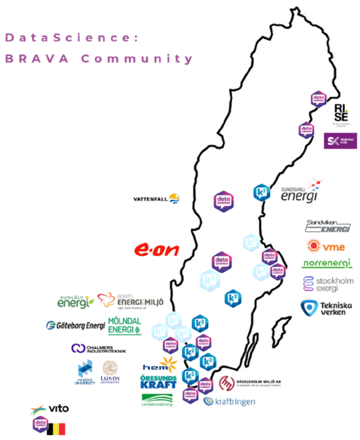
# INDUSTRY COLLABORATION FOR ADVANCED ANALYSIS OF HEAT DISTRIBUTION AND HEATING NEEDS

REPORT 2022:867



FUTUREHEAT




Energiforsk

# Industry collaboration for advanced analysis of heat distribution and heating needs

## Data Science DS:BRAVA – A project for smarter district heating

**FILIP WÄSTBERG**
**MARIA HANSSON**
**RIKARD EDLAND**

ISBN 978-91-7673-867-2 | © Energiforsk maj 2022
Energiforsk AB | Telefon: 08-677 25 30 | E-post: kontakt@energiforsk.se | www.energiforsk.se

# Foreword

**This project started with the aim to increase the analytical capability in the district heating industry. The project has solved various analytical challenges and has showed how district heating companies can collaborate on digital innovation.**

The project was led and conducted by Maria Hansson together with her collegue Filip Wästberg at Solita and Rikard Edland from Chalmers University of Technology.

Fabian Levhin, Stockholm Exergi (chair); 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 and Fredrik Nyman, Kraftringen has followed the project and assured the quality and the usability of the results.

The project is part of the FutureHeat program, with the long-term goal to contribute to the vision of a sustainable heating system with successful companies that utilize new technological opportunities and where the investments made in district heating and cooling are utilized to the best of their ability. This project is part of the second phase of the program.

The FutureHeat program is led by a steering committee consisting of Jonas Cognell, Göteborg Energi (chair); Anders Moritz, Tekniska verken i Linköping; Anna Hinderson, Vattenfall AB; Charlotte Tengborg, E.ON Värme Sverige; Fabian Levihn, Stockholm Exergi; Holger Feurstein, Kraftringen; 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 Hjärtstam, Borås Energi och Miljö; Svante Carlsson, Skellefteå Kraft; Ulf Lindquist, Jämtkraft and Julia Kuylenstierna (co-opt), Energiforsk. Deputies have consisted of Ann Britt Larssson, Tekniska verken i Linköping and Peter Rosenkvist, Gävle Energi.

Julia Kuylenstierna

These are the results and conclusions of a project, which is part of a research programme run by Energiforsk. The authors are responsible for the content.

# Sammanfattning

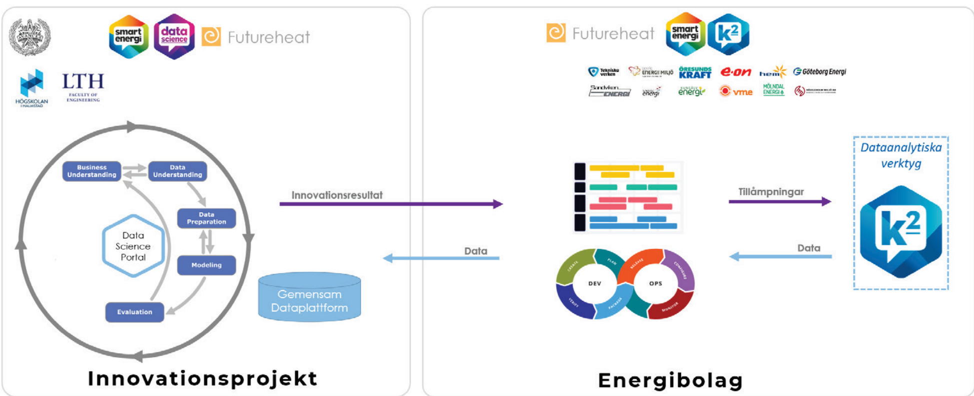

**Genom ett omfattande samarbete mellan fjärrvärmebolag, akademi och data-science- expertis har innovations- och forskningsprojektet DS:BRAVA banat vägen för framtidens smarta fjärrvärme. Syftet är att visa på hur man kan etablera en effektivare och snabbare cykel för att arbeta fram nya och ständigt förbättrade metoder inom avancerad analys- och maskininlärning/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, avvikelsedetektering, taxonomi, smart fjärrvärme, smart energi, machine learning, felkoder, data science, proaktivt underhåll, värmedistribution, labeled data, digital infrastructure

# Summary

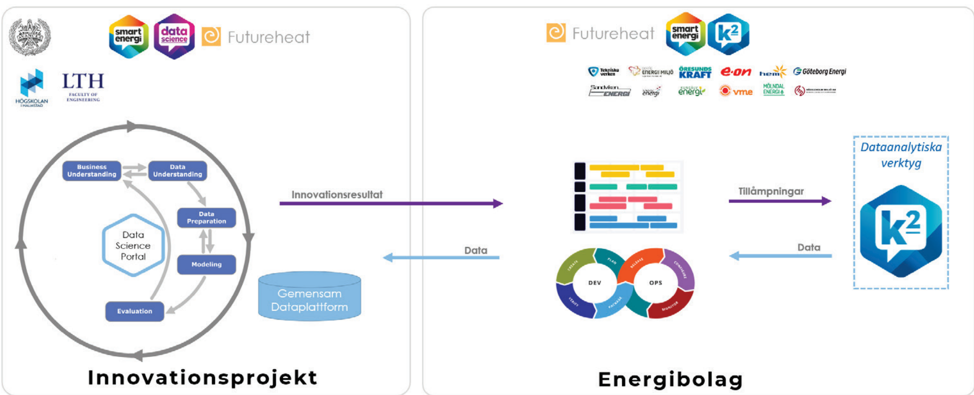
**Through extensive collaboration between district heating companies, academia and data science experts, the innovation and research project DS: BRAVA has paved the way for smart district heating of the future. The purpose is to show how to establish a more efficient and faster cycle to develop new and continuously 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 for 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 within the 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.

# Content

# 1 Introduction and background

According to the report *Fault detection in district heating substations* [5] it is found that 75**% of all district heating substations contain faults that could be detected by analysing deviations**. The deviation analysis can be done manually, but to scale and offer competitive methods, deviation detection needs to be **automated and data analytics methods are a very effective approach to this**.

Identifying deviations in data from district heating is difficult and takes time. Most district heating companies have initiatives to use advanced analysis to better identify deviations and faults in district heating systems. The problems encountered by companies are similar. There is a lack of competence and scalable tools. Above all, there is a lack of data that is needed **industry-wide or standardized labeled data that can be used both in common tools and in research and innovation projects** such as this project.

As district heating distributors face the same or similar challenges, innovation and solutions can with great benefits be shared. In addition, the industry can together increase skills and take joint responsibility for the development needed.

In addition, there are major benefits and economic **gains from working proactively with deviation** and **action planning.** A challenge for district heating companies is to retain customers, since a lost customer means a loss of income that cannot be compared to the artificial reduction it entails from lost customers. District heating production and distribution need to work with proactive optimization to nurture their customer relationships, strengthen competitive advantages and justify the collaborative potential of district heating.

## 1.1 LACK OF HIGH-QUALITY DATA

Statistical models, broadly referred to as AI, are behind, for example, self-driving cars and facial recognition. They are usually developed on data with clear and precise labels (eng. *labeled data*). If we first allow a human to mark up, for example, images of cats and dogs, we can then divide the images into pixels, convert these into numbers and train an algorithm or statistical model that can replicate human sorting. That's what we call machine learning.

Virtually all district heating companies have initiatives to better identify and find deviations in data. One of the biggest problems is the lack of industry-tagged/labeled data, meaning there are not enough deviations that have been registered in a data perspective. The industry is lacking high quality data for innovation and machine learning initiatives.

## 1.2 LACK OF COMPETENCE AND TOOLS

It is possible to move forward even without labeled data/labels. The industry also needs to work on spreading knowledge about data analytics methods, what it means, set requirements, what opportunities it provides, collaborative benefits, etc. The industry needs to increase competence in everything from customer/plant-related roles, to business development, planning, prioritization and management. DS:BRAVA has arranged several public webinars and events where results have been presented, knowledge has been shared and participants have been able to discuss and ask questions.

Furthermore, promoting cooperation between universities, companies and industrial specialists is of great importance. The DS:BRAVA project has focused on making it easy for researchers and specialists from different industries to share code, results and data.

To move forward more quickly in advanced analysis in district heating, from idea to implementation, a common framework for code is required. Companies selling analytics services do not, for obvious reasons, share the code. The DS:BRAVA project therefore focuses on being able to seamlessly disseminate code and analysis results between partners and offer digital infrastructure to enable this framework to be put into practice. Smart Energi provides data from energy companies by sharing data from the K2 application to DS:BRAVA Data Lake. The data sharing takes the form of an Data Sharing API that the Energy Companies themselves can activate. The data is then shared anonymized from each energy company's application K2 to DSBRAVAS Data Lake. DS:BRAVA takes part of the  data and conducts research and produces recommendations for new innovations. Results in the project can then be evaluated and validated in software by further examining how innovation results can be implemented in practice, i.e. in software that energy companies can access. The circle closes.

## 1.3 OBJECTIVES AND PURPOSE

The goal of the project DS:BRAVA is to develop verifiable analysis models based on the energy companies' priority, which can be used to identify leaks, control errors and identify energy patterns. To achieve this, we need a platform where energy analysts, business developers, software developers and data scientists can access aggregated training and test data as well as analysis models.

To enable more advanced data analysis at energy companies, in the form of advanced data analytics and AI models, labeled data is also needed.

Finally, a simple method is needed for companies and organizations to be able to identify the value of new opportunities created from the advanced analysis.

## 1.4 FINANCING AND PROJECT ORGANIZATION

The project is a collaboration between the Swedish Energy companies in the form of the trade collaboration organization Smart Energi, Eneriforsk:Futureheat, The Swedish Energy Agency, Halmstad University, Chalmers Industriteknik. Lund University, Borås University and RISE have also been of great importance for the project and its results. Solita has had project follow-up and coordination responsibilities, as well as contributed with expert resources in Data Science/AI/ML, Data Management and software development. Smart Energi has provided the technical infrastructure for the project in the form of technical environments for Data Lake, Data Science Portal (Kyso), as well as automatic data sharing mechanisms from Smart Energy Energy Company's application K2.

All member companies in Futureheat and Smart Energi are entitled to the project results and can choose how they will be taken into application. For example, individual companies can initiate/order implementation of results in any software or work process, or several companies can co-initiate and co-create results in shared software. The project has not produced recommendations or instructions on how to apply results.

A Data Science council with representatives from both energy companies, expertise in both district heating and data science has continuously contributed with reasoning about priorities and expected value of different elements of the project.

The project is funded by the Swedish Energy Agency TERMO, Energiforsk Futureheat and smart energie cooperation organization. The picture below explains the funding and how participants in the project have been financed.

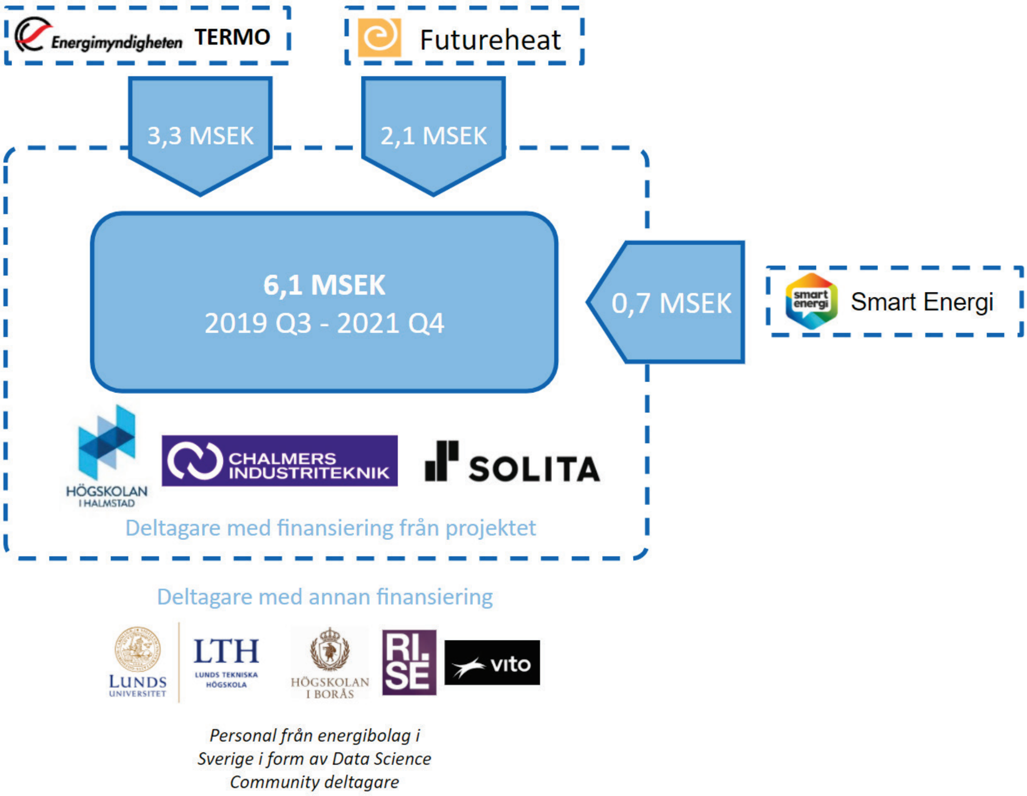



**Figure 1 Schematic picture of the project's financing**

All the people who participated in the project's public webinar and discussions have contributed to the project's results and goal fulfillment.

The image below relates this industry-wide innovation project DS:BRAVA to each Energy Company's own participation, during the project or after the project.

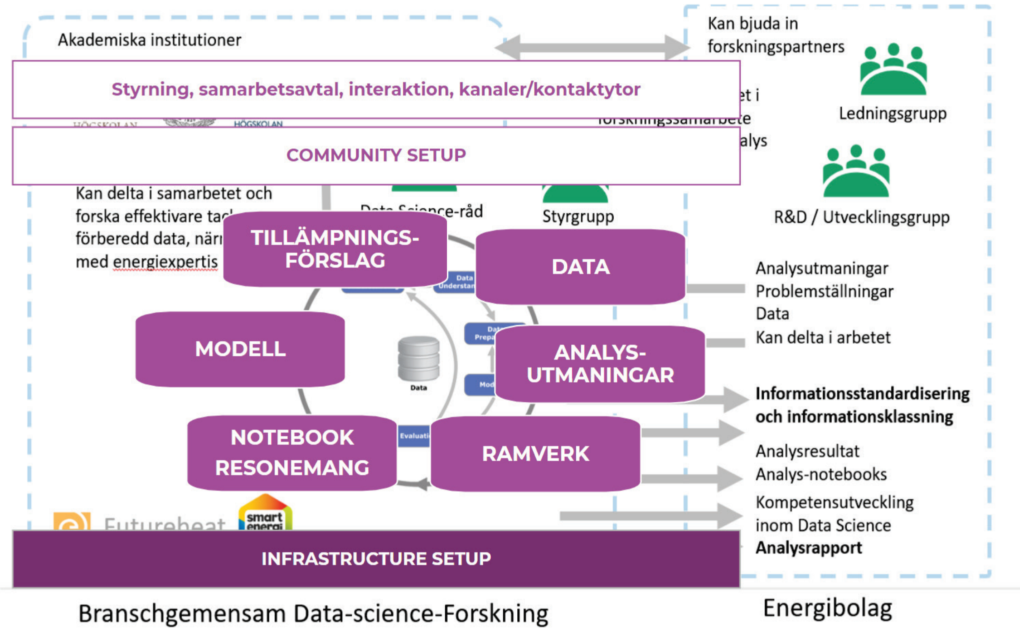



**Figure 2 Project organization in DS: BRAVA**

# 2 DS:BRAVA Data Science-portal

Within the framework of DS:BRAVA, a Data Science portal has been developed to share dataset, reasoning, results, suggestions for analysis challenges and knowledge sharing.

The image below shows how data from the companies is divided into a Data Lake, where analytical reasoning and methods are developed. All reasoning and results are shared in a Results Portal, i.e. the DSBRAVA Data Science Portal.

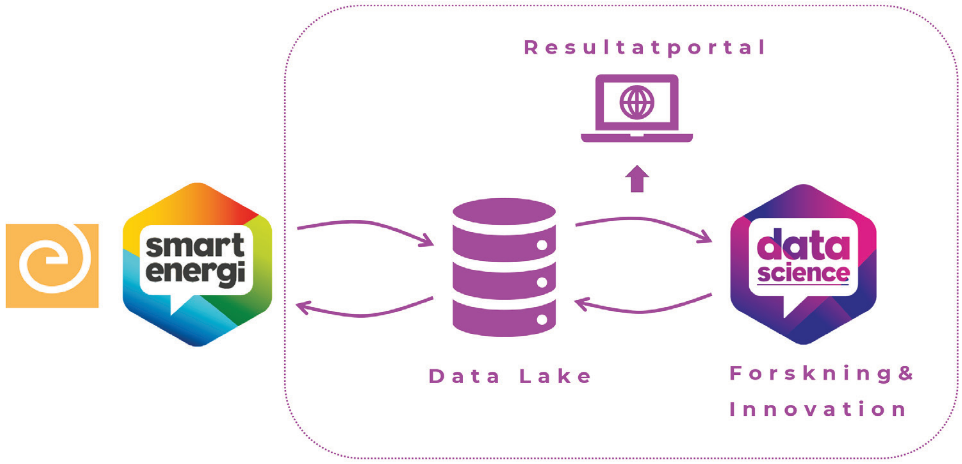



**Figure 3 Data from the companies is shared to a Data Lake**

The content on the Data Science portal consists of dataset, analysis challenges, analytical reasoning, and analytical models. There are currently 37 objects on the Data Science portal that look like the following.

All analytical reasoning and models published on the Data Science portal are also published with associated codes in the Data Science Notebooks. This means that all codes published can be used by members.

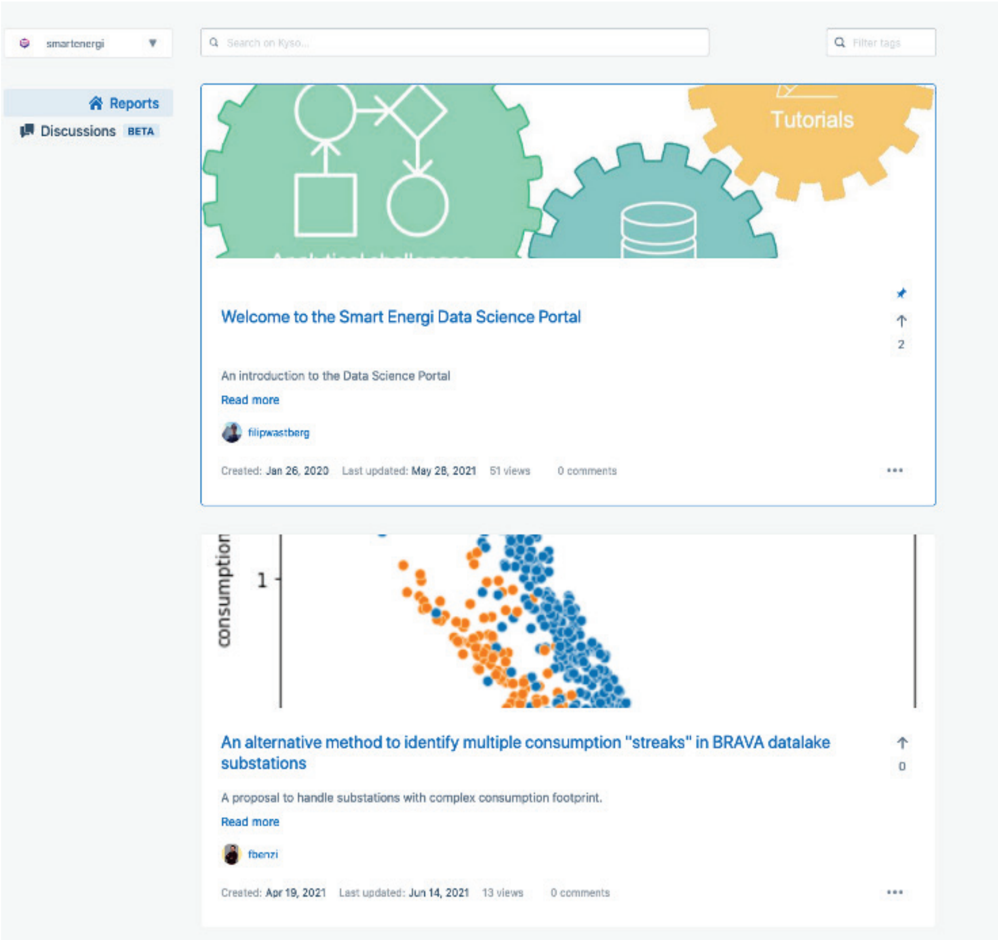



**Figure 4 Data Science portal, home page**

The portal is located at the following address:
https://kyso.io/smartenergi/dashboard?team=smartenergi

Access information: https://smartenergi.org/datascienceDS:BRAVA/

# 3 DS:BRAVA Data Lake

DS:BRAVA has established a Data Lake so that companies that have valuable data for research and innovation purposes can share data with DS:BRAVA. The companies that use K2 are given the opportunity to automatically and anonymously share data contained in K2. Data can also be loaded using manual transfer methods. There are many advantages to automating the data sharing rules because security is automatically followed and because it facilitates the continuous sharing of data from energy companies to innovate projects. Energy companies can choose which data they want to send to DS:BRAVA Data Lake.

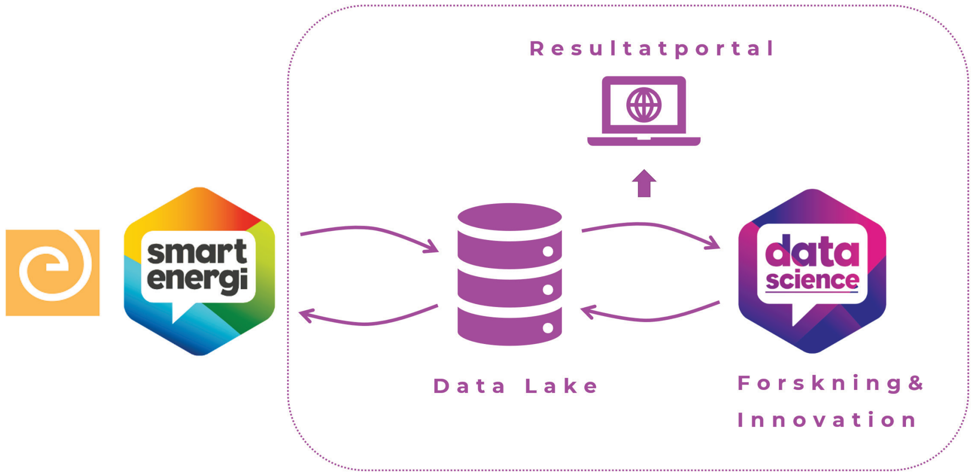



**Figure 5 Data from companies is collected in a so-called Data Lake**

# 4 Deviation Cause Taxonomy

There are a variety of reasons why a deviation occurs in data that comes from district heating customers' heat meters. All these reasons must be easily and efficiently linked to deviations in the data to create tagged datasets. It is also important that as much as possible use the same labels to create a large set of data that can be used to develop better data analysis models. Therefore, a common taxonomy is needed to standardize deviations in district heating data - a deviation cause taxonomy.[6]

There are two main purposes of such a taxonomy: (i) to provide the district heating industry with a structured and standardised way of labelling deviations arising in district heating customer data due to errors in district heating systems, and (ii) to help create labeled datasets for the development of predictive models that can automatically detect faulty district heating systems. The taxonomy should consist of as many different deviation causes as possible, and also include information on when and where in the district heating system the cause of the deviation has arisen. This means that the taxonomy must contain a great deal of information about the different reasons for deviation, and that information must be structured in a way that is logical and easy to understand and follow. Each cause of deviation should be mentioned with a label consisting of a few words clearly describing the cause of deviation. The idea is that this label should be used to label datasets that contain anomalous data patterns, so that it is clear what is behind the discrepancy in the data. With this data, statistical models can then be trained to recognize different deviation patterns.

The easiest way to organize all this information is to arrange it in different levels. The deviation taxonomy developed within this project contains the following levels, all of which will be described in more detail below:

- Cause of Deviation
- Component
- Reason
- Action
- Status

The figure below presents the overall structure of the anomaly taxonomy.

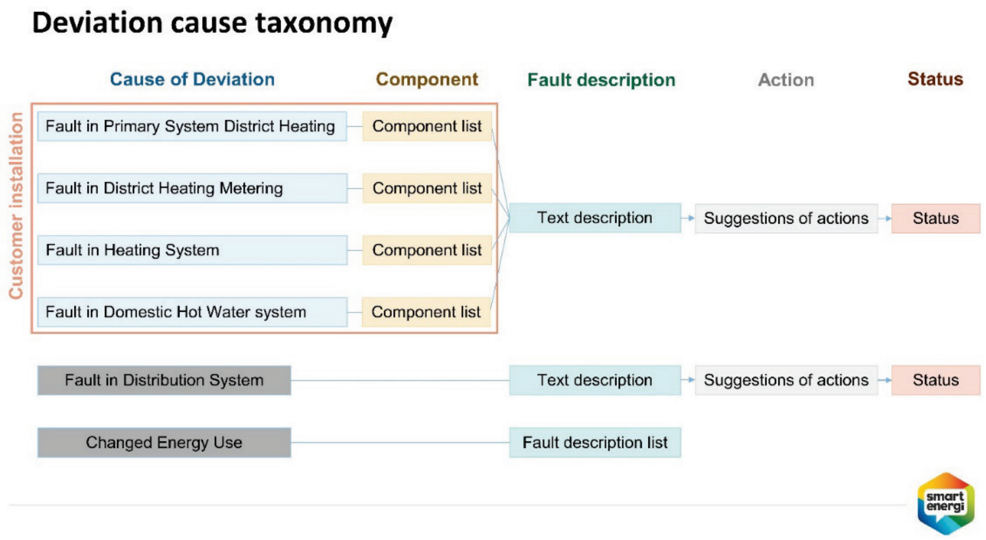



**Figure 6 The overall structure of the cause of deviation taxonomy**

## 4.1 THE FIVE LEVELS OF THE DEVIATION CAUSE TAXONOMY

### 4.1.1 Cause of Deviation

The first level of the taxonomy is *the Cause of Deviation.* This level consists of a number of different categories that describe where the problem that causes the deviation in customer data is located in the district heating system. There are many ways to divide the deviation causes into categories. In this taxonomy, they are divided by function, i.e. in which system or part of a system the causes occur and the effects on the function of said systems. Therefore, the reasons for deviations are divided into the following categories in this taxonomy:

- Faults in Primary Circuit District Heating
- District Heating Component
- The Heating Circuit
- The Domestic Hot Water Circuit
- Changes in energy use
- Distribution Network Error

The first four categories are related to errors that occur in customers' buildings and that cause discrepancies in customer data. This is the categories that contains the most deviations. The orange box in the figure above indicates that these anomaly reason categories are related to errors in the customer's building. The *category Changed energy use* is also related to the customer's building, as this is where a change in energy consumption occurs. However, a change in energy use does not, in many cases, mean that something has gone wrong in a building. For example, a change in energy use can occur if a family with four children who play football five times a week take more showers than a normal single occupant. This will result in a change in energy consumption in the villa, which is likely to appear as an anomaly in customer data. However, this is not an actual *error*, but rather a change that is acceptable and that will persist over time. Therefore, the Changed *Energy*

*Use* category is not directly related to faults in the customer's building, as the four deviation reason categories representing actual faults in the customer's building.

### 4.1.2 Component

The second level of the taxonomy is *Component*. This level describes the individual problem or error that causes a discrepancy in customer data. The four categories related to customers' buildings have at this level a component list that is unique for each reason category. The Changed *Energy Consumption* and Distribution Network Errors categories do not have component lists, but for deviation reason categories. The user should instead directly fill in a Third Level Error *Description* of the taxonomy. There are many components that could cause errors in the distribution network. However, it is likely that many of these errors will not give rise to discrepancies in customer data. Therefore, the taxonomy does not have a component level for the distribution network. Therefore, the idea is that, on the occasional instance when there are errors in the distribution network that affect customer data, the user should fill in information about the error that has occurred in the third level of the taxonomy. Energy change has no component level for the simple reason that it is not relevant to have a component level for this deviation. If there is a problem with a component in the customer's building that gives rise to a deviation in customer data, the idea is that this should be covered by the four deviation reason categories that represent errors in the customer's building.

### 4.1.3 Error Description

In the third level of the taxonomy, the user should describe what is wrong with the component listed at the component level (for the four categories related to the customer's building). An example of this is that during a service visit to a building, a service technician concludes that the deviation in customer data is due to a control valve being broken in the domestic hot water system. In this example, the deviation reason category is *Fault in the Tap Water System,* the component is *Control Valve*, and the fault description would be *Broken.* The idea is also that the user will be provided with several suggestions for error descriptions when this level is filled in, which is based on previous inputs for this level.

For the deviation reason category Changed energy consumption, there is a list of suggestions for different error descriptions. Examples of such fault descriptions are *Changed Heating Requirement, Intentional Shutdown,* and Changed Additional *Heating Supply*. For Distribution Network Errors, the user must fill in a (brief) free text description of the error that has occurred, which has given rise to discrepancies in customer data.

### 4.1.4 Action

The fourth level of the taxonomy is called *Action*. This level represents the action that has been performed to fix the error that has caused the discrepancy in customer data. As a change in energy consumption (usually) is not something that needs to be addressed, this variance category does not have a level of action in the taxonomy. The level consists of several different options of actions:

- Replaced device
- Repaired
- Planned maintenance
- Change of setting
- Etc.

It is also important to specify when an action has been performed, which means that the person using the taxonomy must specify a date and (approximate) time when the action has been performed. This is because it may be interesting for a data analyst to compare the data before and after an operation has been performed, to investigate whether the pattern of the data has changed and whether the deviation caused by the error has disappeared or not.

### 4.1.5 Status

The fifth and final level of the taxonomy describes the status of the reason for deviation, that is, whether or not something has been done to fix the problem. The taxonomy includes three different statuses: *Managed, Partially Managed,* and *Not Managed.* Similar to the Action level, it is important to specify the date and time of the status.

**An example of using the taxonomy**

Before presenting the full component lists, we want to provide the reader with an example of how the taxonomy can be used to record information about an identified error. Let's say that during a service visit, a service technician concludes that the deviation in the customer data has most likely been caused by a control valve that has broken in the domestic hot water circuit and thus no longer works. The service technician also decides to fix the fault directly by replacing the control valve, and then reports back to the district heating company by filling in the information about the fault in a user interface based on the contents of the taxonomy. In this example, the reason for deviation would be "Fault in the Tap Water Circuit", the component is "Control Valve", the error description is "Broken", the action is "Replaced Unit", and the status is "Fixed". Of course, the user interface also fills in the time and date when the action was performed. This example is illustrated in the figure below.

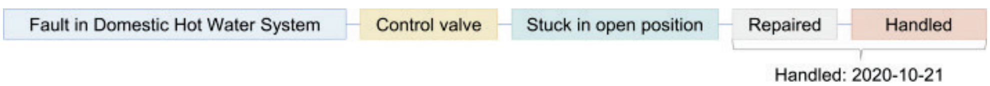



**Figure 7 Example of what a sequence of events might look like**

## 4.2 DETAILED DESCRIPTIONS OF COMPONENT LEVELS

The second level of the taxonomy, Component, contains many components that can give rise to discrepancies in customer data for the four anomaly reason categories related to errors in the customer's building. Below, these four reason categories of deviation, as well as which components are included in the respective lists, are presented. The complete component lists are presented in tables after each individual category has been presented. The tables also contain a number of synonyms for each component collected from the district heating companies. The synonyms are on the same line and are separated by "/". The first word in each line is marked in bold, which indicates that this is the word that most district heating companies surveyed use when naming the different components. For each nonconformance reason category, it should also be possible to specify a free text option in a future user interface if the component that has somehow failed is not included in the different component lists.

### *Faults in Primary Circuit District Heating*

The components belonging to Primary Circuit District Heating are components that are located in the customer's building, but do not belong to any of the other deviation reason categories. What the components have in common is that they are all located on the primary side of the heat exchanger(s). The component list includes, among other things, service valves that turn off the incoming district heating water during maintenance or renovation, hot-holding valves that allow a flow in the district heating system even on hot summer days, filters, and drain valves.

**Table 1 Components that belong to faults in the primary circuit**

| Cause of the Deviation | Component |
|---|---|
| Faults in Primary Circuit District Heating | **Differential Pressure**<br>**Filter**<br>**Pressure gauges** tables /pressure gauges /<br>**Wire**/Pipe<br>**Aeration Valve**<br>**Service Valve**<br>**Swing Valve**<br>**Power Limiting Valve**<br>**Fuse electricity supply**<br>**Drain Valve**<br>**Warm-up service**/Roundabout |

### *The Heat Meter*

The heat meter is connected to the district heating substation and measures the amount of heat consumed by the customer. The meter measures the flow through the district heating centre as well as the forward and return temperature, and then calculates the amount of heat that the customer uses. A communication unit communicates these values to the district heating company's measurement system, after which they are further used for invoicing and measurement value analysis. This means that there are a number of errors that can occur along this transmission chain and that can occur at both the customer and the district heating company. In the table where the complete component list is presented, this is indicated by a dotted line – components listed above this line are in the customer's building, while components below this line are located at the district heating company.

**Table 2 Components belonging to the deviation category Errors in district heating measurement**

| Cause of Deviation | Component |
|---|---|
| District Heating Measurement | **Flow Sensor**<br>**Integration Work**<br>**Communication Unit**<br>**Temperature sensor at front**<br>**Temperature sensor return**<br>**Other disruptive equipment**<br>**Poor communication**<br>**Measured Valve Collection System**<br>**Validation System** |

### *Heating Circuit*

The heating circuit is located inside the customer's building and is the system that delivers heat to each individual room. The circuit can contain a variety of components depending on its design and the heating option installed in each building. In the taxonomy, Heating Circuit Failure contains components on both the primary and secondary sides of the heat exchanger.

**Table 3 Components belonging to the deviation category Fault in heating circuit**

| Cause of Deviation | Component |
|---|---|
| Heating Circuit Failure | **Shut-off valve**<br>**Circulation Pump**<br>**Expansion**<br>**Filter**<br>**Heating**<br>**Ventilation**<br>**Pressure gauge stables/**pressure gauges<br>**Filling Valve**<br>**Radiator**<br>**Radiator vent /**Thermostat<br>**Regulatory Centre/**Regulation<br>**Regulator Valve** /Control Valve /Primary Valve /Control Valve<br>**Shunt Valve**<br>**Actuator**<br>**Safety valve**<br>**Indoor temperature sensors**<br>**Temperature sensor secondary front**<br>**Temperature sensor secondary return**<br>**Outdoor temperature sensors**<br>**Ventilation Unit**<br>**Heat exchanger** |

### *Domestic Hot Water Circuit*

The domestic hot water circuit prepares and delivers hot water to all tap sites in a building. The design of the domestic hot water circuit varies depending on the size and purpose of the building. Therefore, there are a large number of components that can potentially be installed in a domestic hot water circuit.

**Table 4 Components belonging to deviation category Fault in domestic hot water circuit**

| Cause of Deviation | Component |
|---|---|
| Domestic Hot Water Circuit | **Shut off Valve**<br>**Check Valve**<br>**Mixer with mixing valve**<br>**Mixing Valve**<br>**Adjustment valve /** throttle valve<br>**Manometer/**Pressure gauge Control center/Regulation<br>**Control Valve**<br>**Actuator**<br>**Indoor temperature sensors**<br>**Temperature sensor secondary return**<br>**Temperature sensor secondary tolet**<br>**VVC-pump/Circulation Pump**<br>**Heat exchanger** |

## 4.3 IMPLEMENTATION OF NONCONFORMITY DEVIATION TAXONOMY

The goal of an implementation of the deviation cause taxonomy is to be able to enrich energyusestatistics with information on definitive deviation causes, collected on site in the field when a repair or service has been carried out. The long-term goal here is to be able to use these anomaly causes to find statistical connections and be able to develop better theoretical models for finding problems through energy use data.

A natural platform for an implementation of the anomaly cause taxonomy is the application K2. What has been developed so far in K2 is basic functionality and is a basis for future development.

The idea is that information about deviation causes in the future should be collected when a troubleshooting/repair or service is carried out by an energy company. Something similar already exists in many of the companies that store information about service and actions performed, in their billing system. But these are rarely linked to actual energy data.

In K2, all deviations are intimately associated with service protocols and in K2 one or more of them are stored together in protocols that in turn is sorted under the facilities to which they belong.

The core of the system consists of two parts. The first is a specification of what a deviation cause is informationally. This includes definitions for all options allowed under the taxonomy. These are listed in the system as error codes with its associated meanings in plain text (Swedish and English). The second part is a validation function that checks all the reasons for deviation to be added to the system database and ensures that the information is correct. This is important because the information comes from external sources, ultimately from people. Validation enables higher data quality, which is a prerequisite for efficient use of algorithms.

The implemented functionality includes visualizing protocols and associated fault codes for a facility. Below is an example of one.

| 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 åtgärdat |

**Figure 8 Examples of protocols and error codes in a plant.**

Protocols with error codes can be inserted into the system in three different ways. These ways correspond to different levels of integration and provide freedom of choice in how K2 is integrated with external data flows.

- For single protocols, there is a web form in K2 where protocols can be added. Rules in the form ensure that the information added follows set rules.
- If many protocols from many service occasions are added, you can choose to upload this information via a file upload. The following file formats are supported: EXCEL and JSON. The system also ensures that the information added follows the set rules.
- Automated access. In the future, there may be a need to automate the submission of data.

To make it easier for third parties to understand how rules and storage of protocols and deviation reasons work, a reference Excel sheet has been created. The sheet contains forms that demonstrate an interface for entering new deviation causes. The form adds new reasons for a deviation and the sheet serves here as a reference for how the information in the deviation reason is structured. The Excel sheet also allows you to read which input options are allowed, in which combinations, as well as translations into Swedish and English for the different error codes.

As mentioned above, the purpose of integrating error codes into K2 is to be able to develop better theoretical models. An important part here is to be able to mark the data in the energyusagesgraphs that look suspicious and attach descriptive labels to them. The strategy is that algorithms can be trained to see statistical correlations between abnormal measurement data (with labels) and actual causes (deviation codes). Below is an example of a K2 label. The label contains a description as well as metadata.



Energiforsk

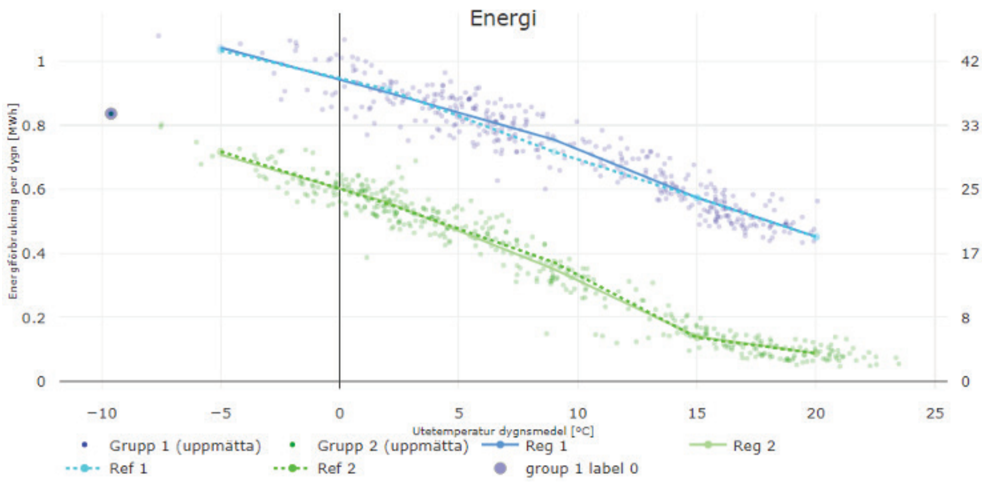



**Figure 9 Example of a label from tool K2**

This list of labels is visualized together with the graphs in K2, where data points with labels are visible. Below is a graph from the same example as above, where the label is indicated (label 0).

# 5 Development of analytical models

One of the main goals of DS:BRAVA is to develop models for identifying deviations in One of the main goals of DS:BRAVA is to develop models for identifying deviations in district heating data. The starting point has been to improve the models used to identify deviations in K2. To identify these deviations, K2 uses data from plants/district heating centres, such as flow, flow temperature and return temperature together with external factors such as weather.

In a classic machine learning scenario, we would have noticed data with many confirmed anomalies, or errors, on which we could "train" an algorithm to identify anomalies and errors in future unknown data. In machine learning, this is called "Supervised Learning", we *monitor* a statistical model or algorithm while learning what is a "correct answer". For example, we can train a model to recognize photos of cats and dogs. But we must then pre-train the mathematical model on images verified by humans containing cats and dogs.

This method is by far the most robust and proven method in machine learning. The problem for DS:BRAVA is that marked district heating data, where we know if something has gone wrong or is a major deviation, does not exist.

The method used by K2 is instead a type of "Unsupervised Learning", where we take advantage of the strong correlation between district heating use and outdoor temperature as shown in Figure 10.

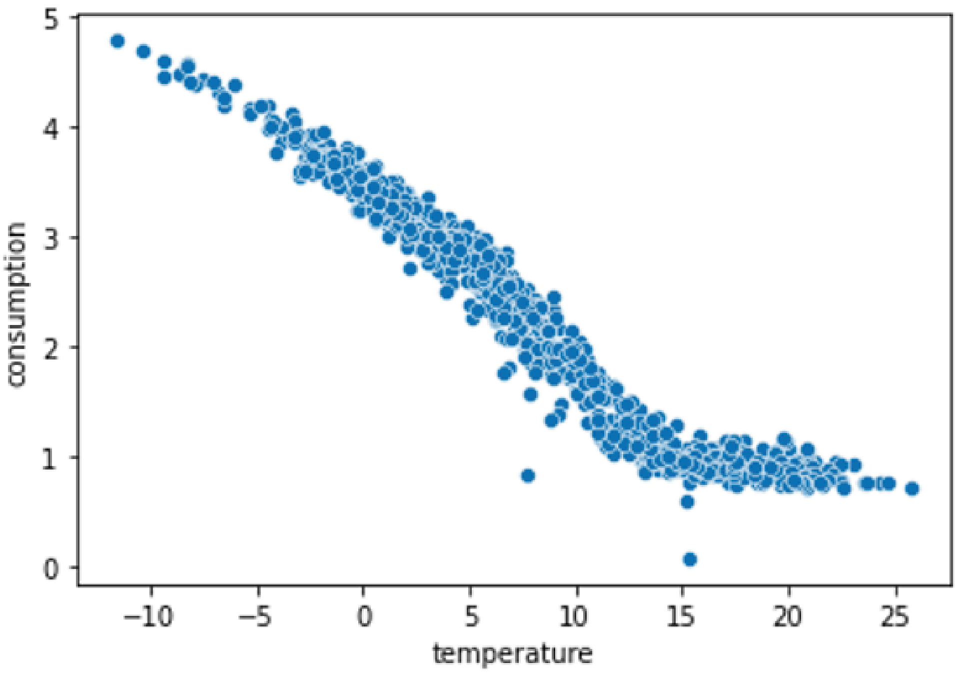



**Figure 10 Correlation between district heating use and outdoor temperature**

Figure 10 above shows the correlation between district heating use and outdoor temperature, which is used in the design of new analytical models and algor timers.

The current algorithm in the analysis platform K2 identifies deviations by per plant/district heating center:

1. Identify a reference period of data without any significant deviations from the energy signature (the relationship between temperature and energy use). The actual identification of a reference period is done by an analyst and business expert.
2. Calculate the relationship between district heating use and outdoor temperature for the reference period with a stepwise linear regression (piecewise linear model). See Figure 10for examples.
3. Use the regression coefficients from the reference period to calculate based on temperatures and the expected energy consumption for the period after the beginning of the reference period.
4. Identify values that are "too far" from their expected values. "Too far" from the expected value is estimated by a standard deviation method.

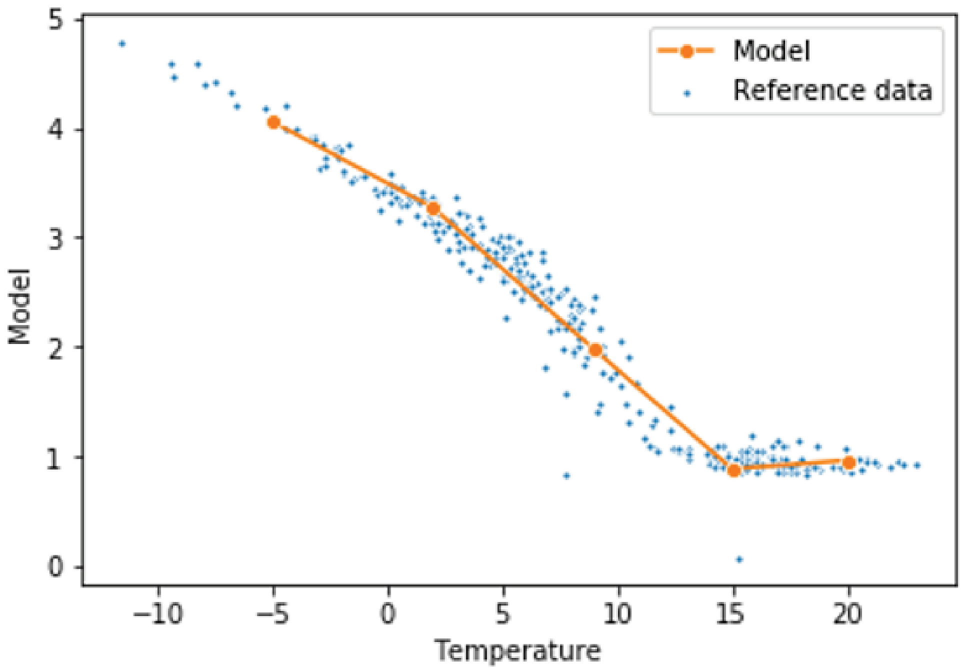



**Figure 11 Reference data and the calculating reference model**

Figure 11 shows the reference data and the calculating reference model for an installation with a step-by-step linear regression.

Details of how the algorithm in K2 works can be found on DS:BRAVA's Data Science portal. The method in K2 has been proven to work well because it is fast and reliable. However, it suffers from some problems.

1. The reference period is more or less arbitrary and needs human support to function well. For a company with tens of thousands of plants/buildings/district heating centres, the method is difficult to use.
2. A step-by-step linear regression is easy to interpret, but there is a risk that the model is not accurate on actual data. This is obvious when energy companies want to use models on hourly data instead of daily data.
3. The method of using standard deviation to identify deviations from expected values is not statistically robust.

In principle, the model in K2 is a forecast model, comparing forecasted values against actual values to identify deviations.

## 5.1 ARNOLD

In the summer of 2020, DS:BRAVA launched an "analytical challenge" where members of the DS:BRAVA project were given the opportunity to compete for which model could "beat" the model in K2. The analysis challenge was described as follows:

> *K2 is an analysis tool for finding deviations in district heating use. K2 calculates the expected district heating use for a location/district heating substation and then compares it with actual use. If the actual value deviates too much from the expected value, an alarm is created. The model currently used in K2 now needs to be developed to better identify deviations. If we can find a model that forecasts district heating use better than K2 does, we will also be able to create more accurate alarms. Most district heating companies use some kind of forecast model for budget and production planning. Can these forecast models also be used to identify deviations more accurately than the current model in K2?*

In total, five people submitted complete answers to the challenge, five models were launched and five models were launched as alternatives to K2's step-by-step regression.

1. A model based on xgboost – a popular model for complex machine learning problems.
2. In time series modeling
3. A segmented regression that is a further development of K2's incremental regression
4. Two contributors used a model called the Generalized Additive Model (GAM) [8]. One of these entries won the competition.

GAMs are popular because, like classical linear regression, they are easy to understand and interpret. However, they have an advantage in that they can be used on data where the relationship between variables is not linear, similar to the relationship between outdoor temperature and district heating use. In Figure 12 you will see a regression line adapted with a GAM.

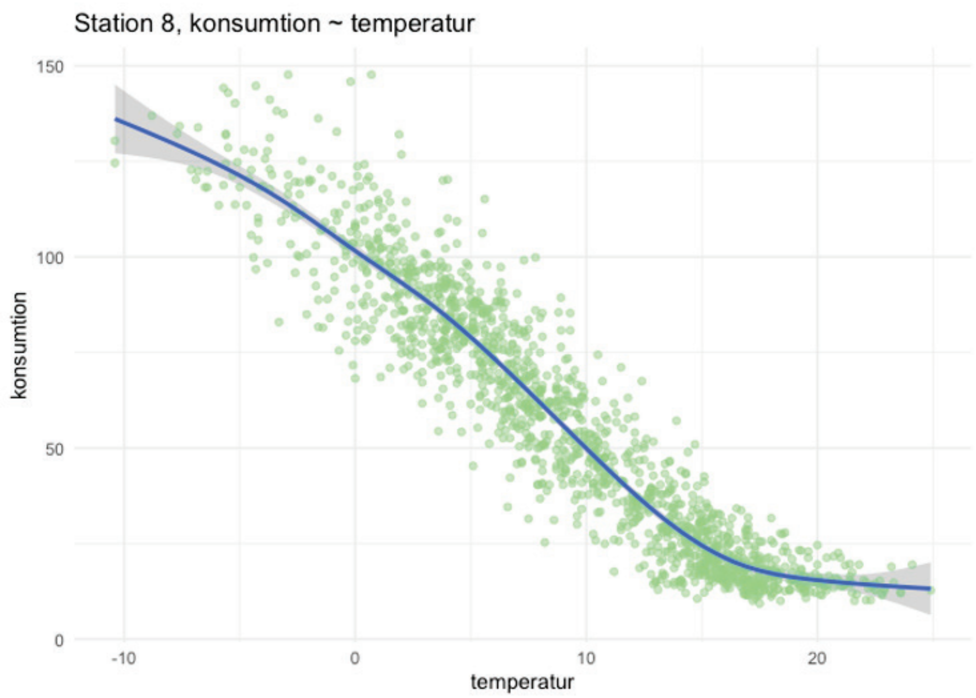



**Figure 12 A regression line between consumption and temperature adapted with the Generalized Additive Model**

The winning entry evolved into a model that could be used in K2 and was named Arnold after mathematician Vladimir Arnold, a Russian mathematician who has been important in the development of GAMs. On the Data Science portal, the idea behind the model Arnold and GAMs is described.

## 5.2 REFERENCE PERIOD

Another problem for the current algorithm in K2 is the choice of reference period. An alternative to K2's reference period selection is under development and uses a method called cross-validation. However, the method is computational, and more work is required to evaluate it.

## 5.3 IDENTIFICATION OF LEAKS IN SERVICE LINES USING CHANGE POINT ANALYSIS

In biomedicine, people are often interested in changes in sequences of various kinds. In district heating analysis, we are primarily interested in changes in time sequences, i.e. when a change occurs over time. In biomedicine there is an area called Change point analysis in which one does just that.

A district heating company asked DS:BRAVA if it would be possible to identify leaks by analysing the temperature of service lines from the district heating system into the house. If a house's energy use, in this case heat use, differs significantly from nearby houses, it should be possible to identify a leak. The district heating company had a data set where a leak had occurred.

In Figure 13 we see the difference between the lead temperature and the return line temperature, delta t, for four villas per month. In the spring of 2014, the delta t drops sharply for one of the houses due to a leak in the service line.

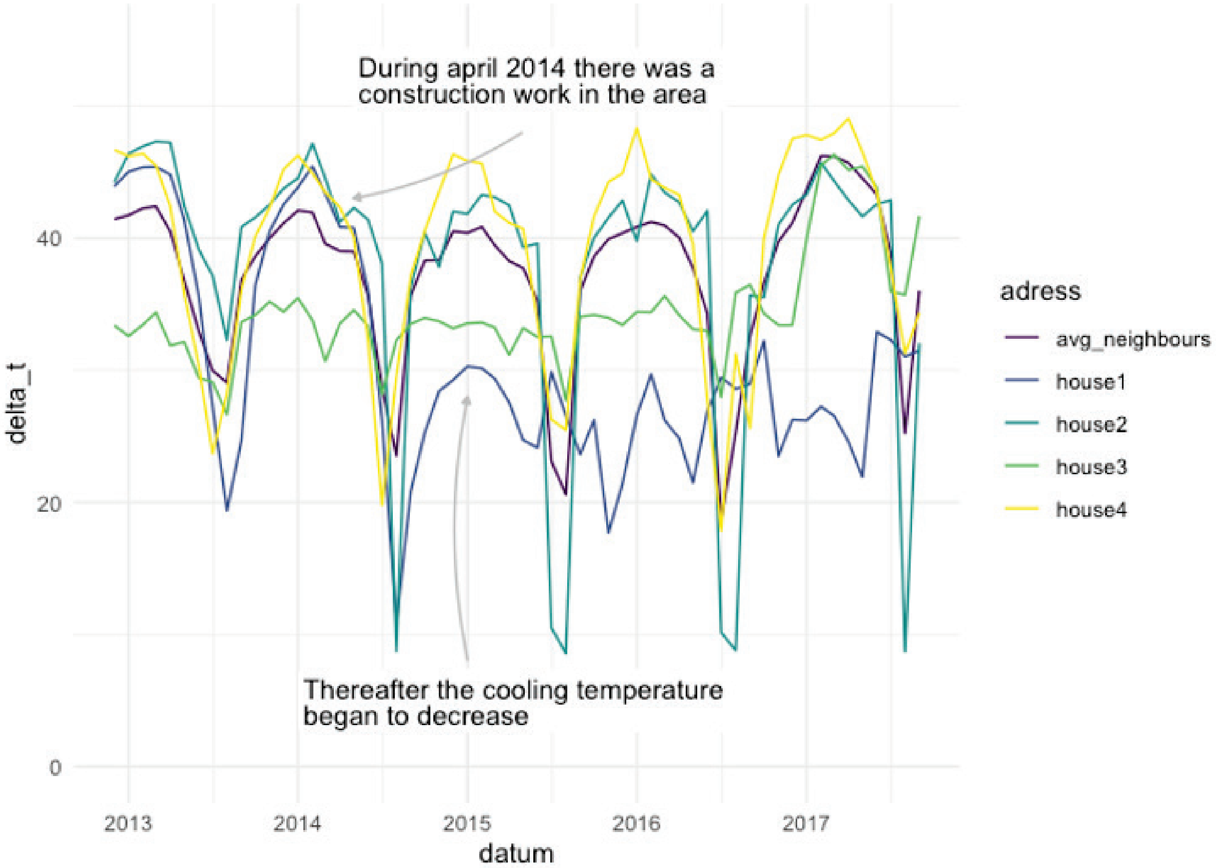



**Figure 13 The difference between flow temperature and return temperature, delta t, for four villas per month.**

A method called Bayesian Changepoint Analysis was applied to this dataset and yielded good results. The method has been evaluated at some plants/district heating substations and works well when there are clear shifts in the data, but worse if they are subtle. Although the method seemed promising, it is difficult to evaluate because data with true leaks is difficult to get over.

### 5.4 VIRTUAL NEIGHBORS

To better identify deviations in data, there was an idea within the DS:BRAVA project that you could look at virtual neighbors to have reference points to analyze against. DS:BRAVA announced an Analytical challenge: Can you find the virtual neighbors?". One of the answers used time series clustering as follows:

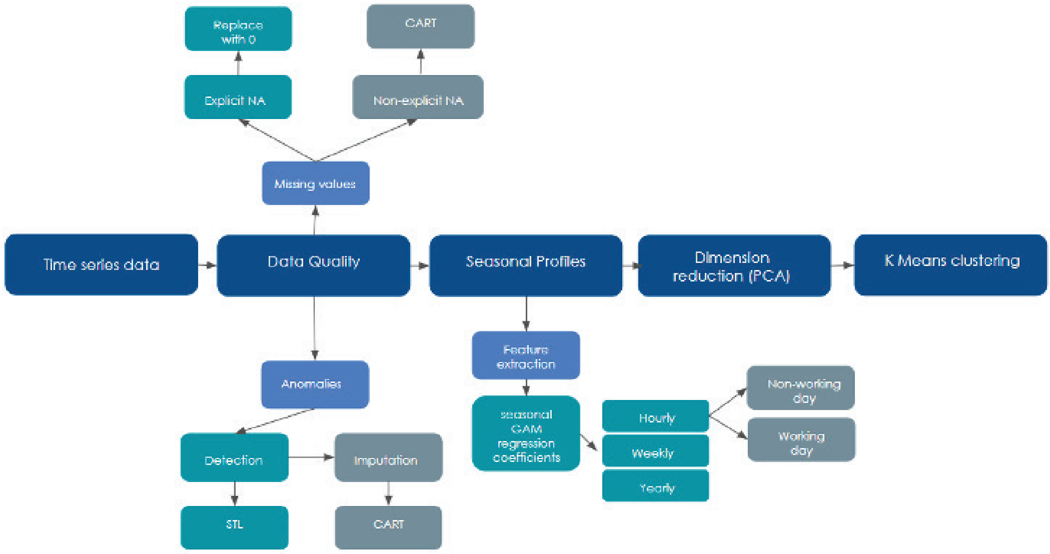



**Figure 14 Time series cluster in the analysis challenge virtual neighbors**

A similar system is used in the models developed by Halmstad University

## 5.5 MODELS FROM HALMSTAD UNIVERSITET

We have developed a data-driven method to enable large-scale analysis of energy patterns in district heating networks and help domain experts understand them better. Our method clusters consumer profiles into different groups, extracts representative patterns and detects unusual consumers whose profile differs greatly from the rest of the group. Using our method, we present the first large-scale comprehensive analysis [3] of energy patterns through a case study on many buildings in six different consumer categories linked to two different district heating networks in Sweden.

We have also developed an anomaly detection algorithm for monitoring district heating systems on a large scale. It is based on an ability to identify a reference group of district heating centres that are similar to a given centre. The reference group is used with a deviation detection framework called "Conformal anomaly detection (CAD)" and has been demonstrated to be effective in detecting several different types of fault categories and atypical energy consumption.

Both works are demonstrated in the form of Jupyter Notebooks in the Data Science Portal.

## 5.6 ANALYSIS OF ENERGY DECLARATIONS

Energy declarations contain interesting information about a household's expected energy use. Within the framework of DS:BRAVA, a large number of energy declarations were analysed. The result was interesting, but exactly how energy declarations should be used in K2, for example, is not obvious.

## 5.7 MULTIPLE ENERGY SIGNATURES

It is not uncommon for district heating data to show energy consumption in several "routes". Often this is due to different usage patterns on different days, for example in an office or school.

Figure 15 provides examples of a "two-stringer" and K2's identification.

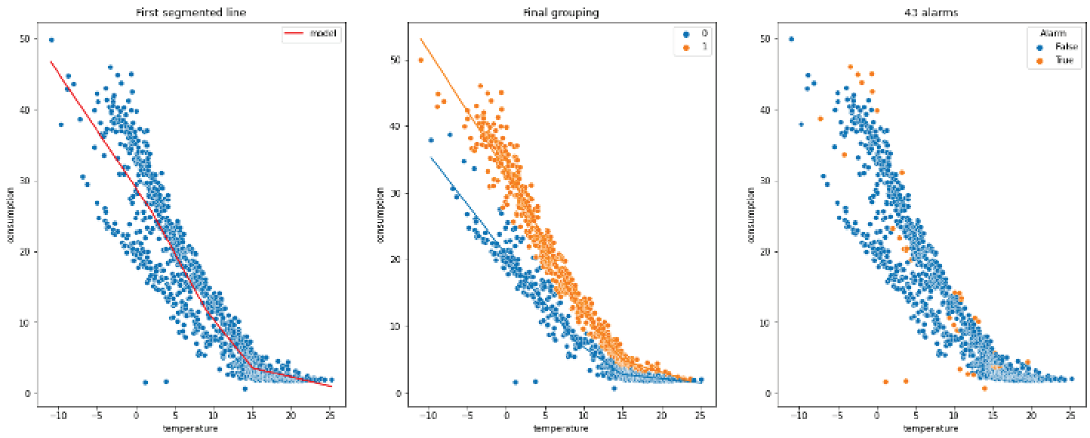



**Figure 15 Examples of identification of so-called " two-stringer "**

At present, K2 identifies paths, which can be two or more, by:

- Calculating a step-by-step linear regression (first graph on the left)
- All points above the line end up in the upper path and vice versa.
- Identify deviations for each group

The method has worked well but has some obvious downsides:

All points in group 0 (in the graph) do not necessarily belong to that group in reality but may as well be group 1 outliers.

The method does not take into account the reason for two-stringers (what day of the week it is).

Within DS:BRAVA, a new method has been developed to identify two stringers where we use the actual cause - difference in weekdays. For example, if we look at the following graph, we can clearly see that district heating use differs between weekdays:

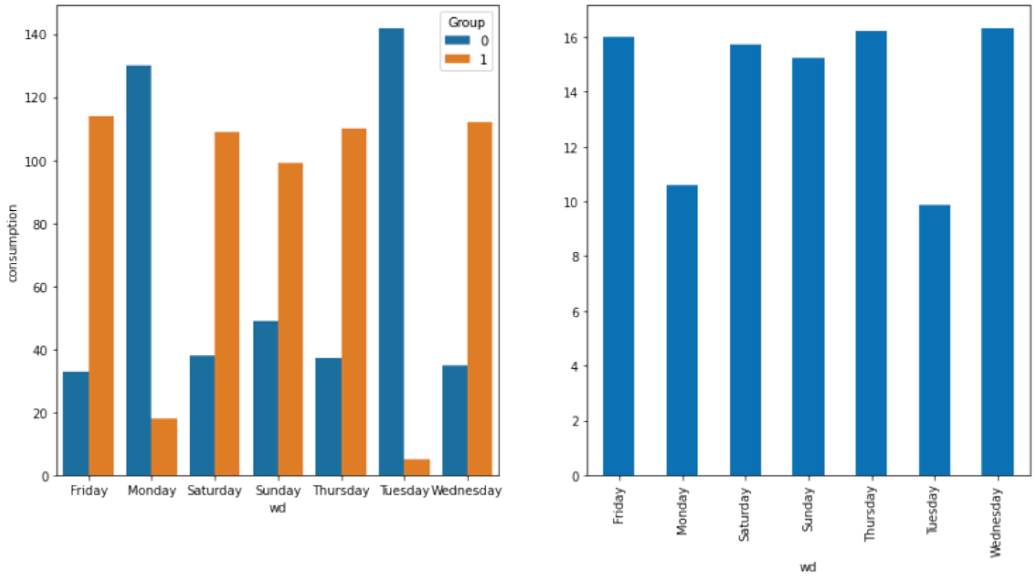



**Figure 16 Variation in energy consumption between different days of the week**

By grouping the observations on the day of the week and using clustering, we can more dynamically find, for example, the two-strings:

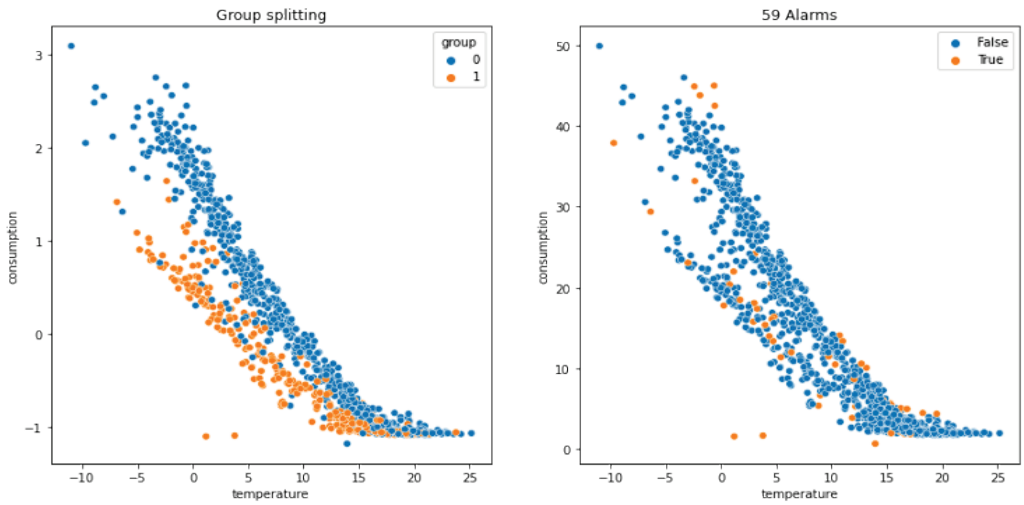



**Figure 17 Examples of how so-called "Two- stringers " can be identified with regard to the variation of energy consumption**

## 5.8 OUTLIERS AND IDENTIFICATION OF ANOMALIES

After adjusting a step-by-step linear regression on district heating data, K2 uses the standard deviation to identify anomalies. Standard deviations work well on data that is normally distributed without outliers. However, district heating data can take many different forms and therefore, within the framework of DS:BRAVA, we have investigated how we can use more robust methods to identify outliers and anomalies. Two of these are IQR, which is one of the most common methods for identifying anomalies in data, and Double MAD, which is a relatively unknown, but effective method, which is robust against different kinds of data. [4]

# 6 Implementation of models in software

## 6.1 SOFTWARE STRINGENCY FRAMEWORK: DS:BRAVALEARN

To enable most analytical models to be applied in parallel, a common framework is needed to write code. In DS:BRAVA, we have therefore developed a module in Python called **DS:BRAVALearn** with which to use models from one of the world's most popular frameworks for machine learning., Sci-kit Learn and pandas (module in Python) to implement models.[9][10]

This framework can be implemented in the analytical applications, software that will be able to receive new analytical methods. This applies, for example, to the K2 application that Smart Energi operates as an industry-wide application, but other applications on the market can apply the same framework. All energy companies that are the clients of this project, DS:BRAVA, can order/ implement DS:BRAVALearn in the tools they use, thus laying a good foundation for new and parallel analytic methods in the future. DS:BRAVALearn is thus an important part of an industry-wide backend for applying advanced data analytics methods.

# 7 Advanced Analysis Value Matrix

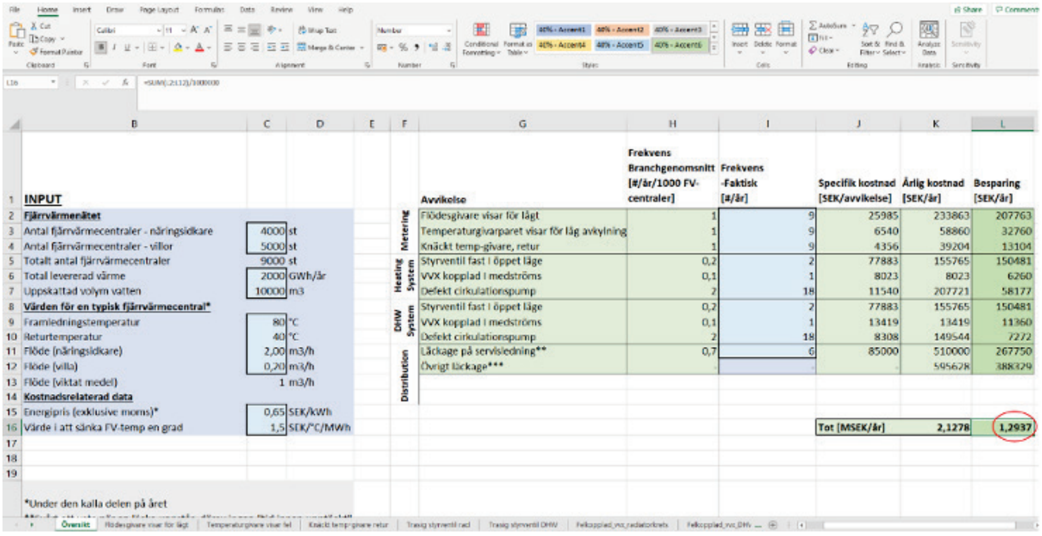
The value of using smart algorithms in the district heating industry is high. Göteborg Energi states, for example, that they have post-invoiced an average of SEK 1.8 million/year (at least: SEK 0.6 million/year, maximum: SEK 4.4 million/year) during the period 2009-2020, largely thanks to smart deviation identification. Some post-crediting also occurs. However, the value of advanced analysis is not limited to post-invoicing. There are additional savings to be made in the form of, for example, reduced return temperatures and reduced leakage.

The development of a tool called the Analysis Value Matrix has started during the project and a preliminary version has been developed. The idea of the Analysis Value Matrix is to provide an overview of the value of identifying (and fixing) specific deviations earlier. A good idea of costs and potential savings can be used for:

- A business case for an energy company to start using advanced analytics
- Prioritize the development of algorithms based on which deviations are most worthy of identifying
- Follow up on the work with deviation identification (in a dashboard, for example)

The methodology in the Analysis Value Matrix to estimate the value of identifying a deviation earlier (for example, that a flow meter measures too low) is to first calculate the cost of a "typical" case where the deviation is identified after a certain time, and then figure out the cost of a case where the deviation is previously identified. The difference between these two cases will be the value of identifying the deviation earlier. Such methodology requires a lot of assumptions and inputs (for example, size of the meter error, time to detection without advanced analysis, time to detection with advanced analysis) which entail uncertainties. However, these assumptions and inputs will be able to be based on better data when more advanced analysis is used and the more data that is collected around different deviations. The accepted inputs in the version that has been developed in the project are based on interviews and estimates/guesses from representatives from energy companies, but there is still a high level of uncertainty.

The annual saving depends partly on the specific value of identifying and correcting each deviation earlier (SEK/deviation) and partly on the frequency of deviations (deviations/year). If an energy company has knowledge of the frequency of a deviation, it can be entered directly into the Analysis Value Matrix. Otherwise, an industry-average frequency has been developed for each deviation (deviations/year/1000 district heating centres) that can be used to estimate the frequency of deviations. This industry-average frequency is currently based on qualified guesses from some of the Energy Companies' representatives but will be able to be more accurate the more advanced analysis is used.



| | | | |
|---|---|---|---|
| 1 | **INPUT** | | |
| 2 | **Fjärrvärmenätet** | | |
| 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 | m3 |
| 8 | **Värden för en typisk fjärrvärmecentral*** | | |
| 9 | Framledningstemperatur | 80 | °C |
| 10 | Returtemperatur | 40 | °C |
| 11 | Flöde (näringsidkare) | 2,00 | m3/h |
| 12 | Flöde (villa) | 0,20 | m3/h |
| 13 | Flöde (viktat medel) | 1 | m3/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 |

*Under den kalla delen på året

| | Avvikelse | Frekvens Branchgenomsnitt [#/år/1000 FV-centraler] | Frekvens -Faktisk [#/år] | Specifik kostnad [SEK/avvikelse] | Årlig kostnad [SEK/år] | Besparing [SEK/år] |
|---|---|---|---|---|---|---|
| Metering | Flödesgivare visar för lågt | 1 | 9 | 25985 | 233863 | 207763 |
| | Temperaturgivarparet visar för låg avkylning | 1 | 9 | 6540 | 58860 | 32760 |
| | Knäckt temp-givare, retur | 1 | 9 | 4356 | 39204 | 13104 |
| Heating System | Styrventil fast I öppet läge | 0,2 | 2 | 77883 | 155765 | 150481 |
| | VVX kopplad I medströms | 0,1 | 1 | 8023 | 8023 | 6260 |
| | Defekt cirkulationspump | 2 | 18 | 11540 | 207721 | 58177 |
| DHW System | Styrventil fast I öppet läge | 0,2 | 2 | 77883 | 155765 | 150481 |
| | VVX kopplad I medströms | 0,1 | 1 | 13419 | 13419 | 11360 |
| | Defekt cirkulationspump | 2 | 18 | 8308 | 149544 | 7272 |
| Distribution | Läckage på servisledning** | 0,7 | 6 | 85000 | 510000 | 267750 |
| | Övrigt läckage*** | | | | 595628 | 388329 |
| | | | | Tot [MSEK/år] | 2,1278 | 1,2937 |

**Figure 18 Image of analysis value matrix in excel**

An example can be seen in the figure above in the form of a screenshot of the first sheet of the Analysis Value Matrix. The tool is Excel-based and in the figure fictional figures have been used for an energy company with 9000 district heating substations (4000 traders and 5000 homes), a network volume of 10,000 m3, forward temperature of 80°C, and return temperature of 40°C. The energy companies can fill in this type of input themselves. In this case, the annual costs of the deviations stated are estimated to amount to just over SEK 2.1 million/year, and the savings potential of detecting the measures previously is estimated at approximately SEK 1.3 million/year. In the example, the deviations with the highest saving potential are: leakage (656 kSEK/year), flow sensors showing too low (208 kSEK/year), and control valve that is stuck in open position (300 kSEK/year). These figures are examples of results that the Analysis Value Matrix can provide, even if these figures have a high level of uncertainty attached to them.

As mentioned earlier, the more data collected, the more data the analysis value matrix can be developed. The current version calculates costs and savings through a bottom-up approach, i.e. by assumptions. In a future where advanced analysis is used to a greater degree than today, the savings can be calculated "top-down", i.e. based on data.

# 8 Community sharing and public webinars / events

An important part of the DS:BRAVA project has been to share knowledge within the industry. This has been done on the Data Science portal but also via open webinars (lunch sessions), workshops in Energiforsk directed and focus days organized by DS:BRAVA project teams. During these events, different people have shared, for example, a new model, avication-causal taxonomy and more.

A total of 2 public conferences/focus days have been organized (including the results website, 11 open webinars to serve continuous project results, 1 editorial article in Energi magazine,2 public workshops where Energiforsk has been convened).

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 Hägskola - *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 avvikelsedetektering
DS-Community: Public Session #10 Sara Månssson disputationspresentation
DS-Community: Redaktionell artikel om samverksansprojekt inom datanaalys för energibranchen - 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

The Community that is being built, i.e. those who have registered as participants in events or applied for eligibility in the Data Science Portal, comes from both energy companies and research institutes and universities.


Figure 19 The image above describes the Data Science Community and its coverage.

# 9 Outlook for continued activities

The project has developed an infrastructure and working methods for a streamlined data science life cycle. It is a value and assets that will be wasted if it is not used and managed in the future. In addition, several areas have been identified as suitable for further work.

These areas **cover a wide** range of possibilities. Some areas are about implementing results from the project, such as energy companies starting to use the taxonomy or implementing the Arnold analytical model. Other areas focus on clearing obstacles to the use of advanced analytics, such as developing a data sharing framework or descriptive data sharing materials and GDPR.

## 9.1 SUGGESTIONS FOR FOLLOW-UP ACTIVITIES

The following table provides an overview of the different tasks with suggested owners. A brief description of each activity follows, as well as suggestions for possible public funding.

| Activity | Suggested owner |
|---|---|
| Data science portal and data lake management | Smart Energy |
| Skills-enhancing activities in the industry | Smart Energy everything. The energy phöretag, everything. Energy research |
| Data sharing framework for research purposes | Smart Energy everything. Energiföretagen, Samverksansdrivet Innovationsprojekt* |
| Management of the taxonomy | Energy companies, alt. Joint venture-driven Innovation Project* to land ownership |
| Management of the analytical value matrix | Energy companies, alt. Joint venture-driven Innovation Project* to land ownership |
| Workflow for detecting anomalies | Smart Energy everything. Energy Companies |
| Application of advanced analytical models | Smart Energy and respectively. Energy companies |
| Leak detection + test bed | Research institutes or universities, Samverksansdrivet Innovationsprojekt* |
| Link between energy companies and real estate companies | Research institutes or universities, Energy Company innovation projects* |

*) This may be a follow-up project of Data Science:DS:BRAVA, e.g. Stage 2.

### 9.1.1 Management of Data Science Portal and Data Lake

The project's interim results and collected anonymized data can be found on the Data Science Portal and Data Lake. How the platforms will be owned, managed and how access should be managed are important issues in the future.

### 9.1.2 Skills-enhancing activities in the industry

The project has built up a national community where many have participated in continuous webinars and activities with the aim of sharing results from the project but above all with the aim of increasing competence in the industry, which was one of the project's goals. This community and actuation will continue.

### 9.1.3 Data sharing for research purposes

Data sharing has proven to be a moment where there is great uncertainty about what can be shared and how it should be carried out. A data sharing framework that parties can agree on as well as a descriptive material about GDPR about what is allowed/sensitive to share could remove many obstacles and speed up the development and application of advanced analysis.

### 9.1.4 Management of the taxonomy

Management and ownership are important for the continued use of the taxonomy. How could it be followed up and possibly improved/adapted, and who is responsible for it? How will the spread of taxonomy in the industry be possible?

### 9.1.5 Analysis value matrix and dashboard

How does the analysis value matrix work as a method for prioritizing and visualizing the composite value of proactive data analysis work? What needs to be developed and how can companies start using this? A prototype dashboard has been created within the project where deviations and costs/values are compiled and visualized. How does it work and how should it develop?

### 9.1.6 Workflow for detecting anomalies

One basis for using advanced analysis is that good data is available. A structured workflow that results in identified deviations being analyzed and labeled is thus a prerequisite for the efficient utilization of advanced analytics. Who is responsible for the way of working and makes sure that it is used?

### 9.1.7 Application of advanced analytical models

Several models have been developed in the project that have not yet been implemented. Arnold and the identification of two stringers are two examples mentioned earlier in the report that can be implemented in the future. Double MAD is another algorithm that has been developed but not implemented.

### 9.1.8 Leak detection + test bed

The value of identifying leaks using advanced analysis is estimated to be high. For example, several energy companies indicate that they have a turnover of approximately one network volume per year, which can correspond to thousands of cubic meters of urban water that need to be supplied, treated and heated per year. In addition, leakage can cause additional damage and also lead to the need to excavate wires. The *Section Identification of leaks in service lines using Change Point Analysis* above can be read about. An example where advanced analysis could be used to identify a leak that occurs. It would be of great value with similar case studies. For example, to test algorithms at district heating centres located in areas where leakage has been identified using aerial identifiers. It should be mentioned that there have been cases where neither algorithms nor experts can see that there is a leak just by analyzing the data, despite knowing that the data set shows values for a district heating centre where there has been a leak on the service line. A central problem is that you rarely know when a leak occurs. In order to obtain better facitdata, a test bed where you can simulate the operation of a district heating centre and leakage would be of great value. Lack of data and facitdata is often a showstopper for innovation.

### 9.1.9 Link between energy companies and real estate companies

Some real estate companies offer their customers smart services, for example in the form of overviews of energy use and other data (water consumption, room temperature, etc.) that are collected via sensors and that can be accessed through a property portal. Energy companies often offer their customers similar services, but could get more involved in how the heat is used. Here is an opportunity to investigate whether energy companies and real estate companies can work together to find ways to steer (or "nudge") energy use to a smarter/more efficient use pattern that benefits both parties.

The transition of the energy system towards an increased amount of real estate energy production in the form of solar cells, air heat exchangers and geothermal heat leads to consumers becoming procumbent with new net consumption patterns with new types of weather (hot or cold and windy and sunny or still and cloudy weather). Energy companies need to be better at building forecasts and control models to control and **balance** the energy system in different weather conditions. There is great potential to create economic and sustainability values through better optimization and co-control of the entire energy ecosystem from production to consumption. A prerequisite for this is that energy companies and real estate companies start sharing data and information in both directions in a structured way and that governance and price models are established for balancing towards more efficient and sustainable energy management. A special opportunity exists here to collaborate between the municipal housing companies of The Public Utility and the municipal energy companies as they are both part of our municipalities and cities. Here there is a challenge to design and establish an operational digital collaboration model based on data exchange, forecast and control models and industry API that enables the digital collaboration model.

## 9.2 POTENTIAL PUBLIC FUNDING FOR FURTHER STEPS

Digitalisation and energy are two areas of technology that are invested in a lot. Below are suggestions for public programmes/periodic calls that could suit the activities mentioned so far.

**ERA-NET SES (Smart Energy Systems).** ERA-NET is one of the European Commission's instruments for research and innovation. The ERA-NET SES initiative works with four areas: smart power grids, integrated regional energy systems, flexible heating and cooling systems, and smart services. Funding goes to transnational projects. The latest ERA-NET SES called for at least three independent parties from at least two different countries. Examples of activities could be to spread the taxonomy and workflow for identifying deviations outside of Sweden.

**THERMO.** The Swedish Energy Agency's research and innovation programme TERMO has financed the DS:BRAVA project with 54% of the total amount. If more calls are released within the program or if there is a follow-up, it would be a natural place to search for continuation activities.

**AI in the service of the climate**. This is a call from Vinnova, Formas, the Swedish Energy Agency and the Swedish Space Agency with a focus on reducing greenhouse gases that have been released in two batches. This call therefore only fits projects that can be expected to lead to reductions in greenhouse gases. Since the district heating industry aims to be fossil-free by 2030, it can be difficult to argue that digitalisation gives rise to large emission reductions. On the other hand, it is not impossible that in the future there will be an increased focus on being resource efficient with the use of biomass as fuel, and then digitization of the district heating sector (and thus an expected reduction in return temperature and a more efficient system) may be relevant for this or similar cases.

**Energy Research / Future heat.** Provided that Energiforsk comes up with further calls.

# 10 Reference List


[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

# INDUSTRY COLLABORATION FOR ADVANCED ANALYSIS OF HEAT DISTRIBUTION AND HEATING NEEDS

The goal of the project has been to increase the analytical ability in the district heating sector. Also, how more efficient and faster cycles to develop methods whithin advanced data analytics, machine learning and artificial intelligence can be established. The project has both responded to various analytical challenges from the industry and also investigated how the industry can collaborate within digital innovation.

Within the project, a taxonomy for labeling data has been developed. The taxonomy creates an homogenization in the industry which makes it easier to manage big data sets and train algoritms.

Data has been collected from district heating companies, primarily from the analysis application K2 to a common data platform connected to a data science portal.