Proje MWh

Kari Luonteri

Projektin tarkoitus .... **Error! Bookmark not defined.**

Ongelman kuvaus .... **Error! Bookmark not defined.**

Miten lähestyä ongelmaa .... **Error! Bookmark not defined.**

Asuntojen lämmitykseen kuluva energia. .... **Error! Bookmark not defined.**

Lämpimän veden kierrätyksessä tapahtuva lämpöhäviö. .... **Error! Bookmark not defined.**

Käyttöveden lämmitykseen kuluva energia... **Error! Bookmark not defined.**

Ongelman ratkaisu .... **Error! Bookmark not defined.**

Patteriverkoston lämpöhäviöt ja niiden tarjoama vuotuinen säästömahdollisuus....... **Error! Bookmark not defined.**

Käytännön sovellus ongelman ratkaisusta .... **Error! Bookmark not defined.**

Lämpimän kiertoveden verkoston lämpöhäviöt ja niiden tarjoama vuotuinen säästömahdollisuus 6

Project Purpose

The project is a practical assignment as part of the "Data Processing and Utilization of Machine Learning" course, organized by Taitotalo as part of a partial qualification.
In this assignment, the student describes a problem that can be solved using data, processes existing data, and solves the problem using machine learning.

Problem Description

In 2022, the total energy consumption in Finland was approximately 294 TWh. Of this, about 27%, or nearly 80 TWh, was used for building heating. Energy consumption for residential use accounts for approximately 60 TWh annually. Out of this, about 43 TWh is used for heating residential spaces, and about 10 TWh is used for heating domestic water.

In residential buildings, hot domestic water circulates in the piping system to avoid unnecessary water wastage while waiting for hot water at the tap. Since this can also reduce the amount of water used, the recirculation of hot water is not entirely wasteful. However, the water circulation in the pipes causes heat losses, which can be around 10-20% of the energy used for heating the domestic water. In buildings where space heating is done using water circulation, a portion of the energy used for heating is also lost in a similar manner.

Heat losses and the energy consumed by them can likely be reduced if the water circulation is stopped when it is not needed. Thus, the problem can be formulated as follows:

CAN MACHINE LEARNING BE UTILIZED TO ANALYZE AVAILABLE DATA TO DETERMINE THE TIMES WHEN THE CIRCULATION OF DOMESTIC WATER OR HEATING WATER COULD BE REASONABLY STOPPED, OR PROVIDE INFORMATION THAT CAN BE USED TO CONTROL THE CIRCULATION PUMPS?

How to approach the problem

The problem is twofold, as water is circulated in two different piping systems: 1) hot domestic water pipes and 2) the radiator network. The first thing to investigate is when energy is consumed in these systems. The energy required for heating domestic water, can be estimated to be relatively constant throughout the year, although it strongly correlates with the amount of water used in the residence. Water consumption varies according to the time of day and the day of the week. There is also some variation between months. This question could be studied using a time series analysis.

On the other hand, the energy used for heating depends on the outdoor temperature, presumably in a linear manner. This can be confirmed, for instance, through linear regression. As an example, we will study data collected from one housing cooperative, which includes water consumption and district heating consumption as a function of time. The data has been collected hourly. The outdoor temperature can be obtained, for example, from the Finnish Meteorological Institute.

The energy used for heating residential buildings

The energy used for heating residential buildings is consumed as radiant heat through radiators, with some being lost as heat during transmission. If the same energy source is used to heat both residential spaces and domestic hot water, this means that it is necessary to distinguish between 1) the energy used for heating the living spaces and the energy used to heat the domestic water, and 2) the portion of energy lost as heat. The energy used to heat domestic water can be excluded from the equation by using data where water consumption is zero or where it can be reasonably assumed that only cold water is being used, as in this case, no energy is consumed for heating water. A linear equation for residential heating can be determined, for example, using linear regression (y = ax + b). If we assume that the portion of energy lost as heat is relatively constant under all conditions, it can be removed from the analysis by examining 1) data where water consumption is zero (or only cold water is used) and 2) cases where no energy is consumed for heating the building.

Heat loss in hot water circulation.

When water is circulated in domestic water piping and radiator networks, heat loss occurs as a result. The magnitude of this can be estimated (or calculated), for example, by examining the minimum energy consumption when two conditions are met: 1) no hot water is being used, and 2) no energy is consumed for heating residential buildings. It should be noted that heat losses occur in both the radiator network and the domestic water piping. It can be estimated that the heat losses are evenly distributed between the two piping systems.

Energy Consumption for Heating Domestic Water

Heating domestic water consumes energy in two ways: 1) Heating the water used, and 2) Heat loss caused by recirculating the hot water. Heat losses can be estimated with reasonable accuracy, as previously mentioned, and by using the total energy consumption, heat losses, and the energy used for heating residential buildings, it is possible to calculate how much energy is consumed in heating domestic water.

The energy consumed in heating domestic water can be calculated with a simple subtraction: Total energy consumption - Energy used for heating residential buildings - Energy from heat losses.

It is of course possible to calculate energy consumption, but in order to eliminate the heat losses caused by recirculating hot water, it would require stopping the circulation of hot water. This, in turn, would require reliably predicting the hours during which only cold water is used in the building, or no water is used at all. The problem is that if the prediction is wrong, water will have to be run, increasing water consumption.

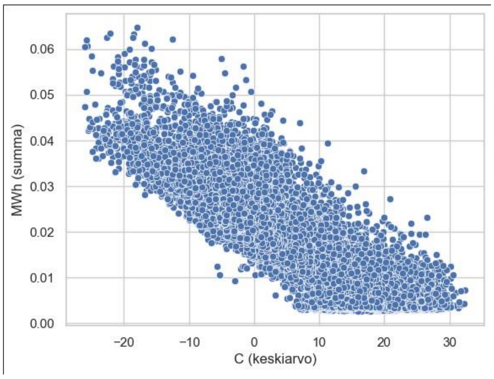
For stopping the circulation of hot water to be worthwhile, it would be necessary to predict with sufficient accuracy the hours during which water consumption is zero or below 10, perhaps 20 liters per hour.

Solution to the problem

I considered it best to divide the problem solution into two parts, for which separate solutions were sought. This makes sense also because the piping for domestic water and the heating network are separate, and their actuators (i.e., pumps) are also separate.

Heat Losses of the District Heating Network and Their Annual Savings Potential

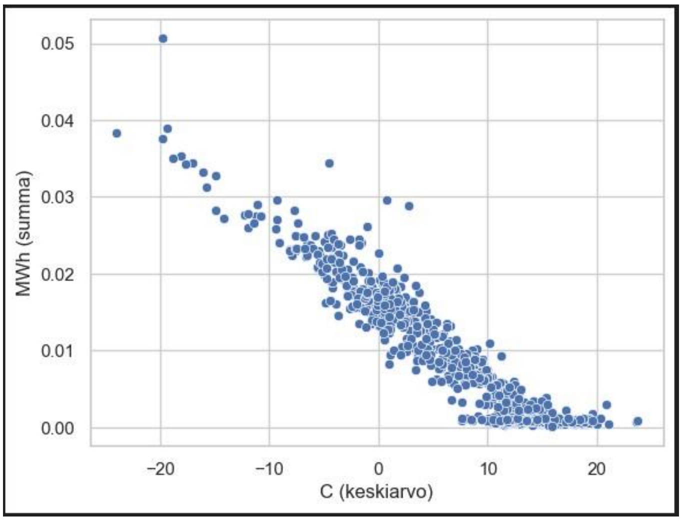
The energy consumption required for heating homes is somewhat directly proportional to the outside temperature, so it is reasonable to use linear regression to determine the equation for energy consumption. Linearity can be easily observed, for example, through a scatter plot.



It is possible that other external variables, in addition to temperature, also affect energy consumption. Such factors may include wind. Since the meteorological data includes wind strength, we will include it in the calculations as well.

The scatter plot on the right shows the total energy consumption compared to the outdoor temperature, including both heat losses and the energy consumption required for heating domestic water. To calculate the energy consumption required for heating the apartments, these two components need to be removed. The energy required for heating domestic water can be easily eliminated by removing data points where hot water is used. This can be done, for example, by removing observations where the hourly consumption is greater than 0.0 m^3/h.

In addition to this, heat losses are also removed by subtracting the minimum energy consumption, which is approximately 0.0025 MWh/h, from the energy consumption of the remaining observations. This results in a significantly cleaner scatter plot on the right.



With this data, the linear regression yields the following equation (there is slight variation in different time periods)

Tällä datalla laskettuna lineaarinen regressio antaa tuloksena (eri ajokerroissa esiintyy pientä vaihtelua) yhtälön:

y = -0.001009x1 + 0.000596x2 + 0.014410,

where *y* is the energy consumption, x1 is the outdoor temperature, x2 is the wind speed, and the constant term is 0.014729. The threshold temperature at which the energy consumption for heating residential buildings drops to zero can be easily calculated from this equation by setting the wind speed and energy consumption to zero. This threshold temperature is approximately 14.3 °C.

The interpretation of this threshold temperature is that when the outdoor temperature exceeds this value, there should be no energy consumption for heating the apartments. In reality, there are some deviations from this, which can also be seen in the scatter plot, but these cases likely relate to variations in heat losses, as the amount of energy involved is very minimal.

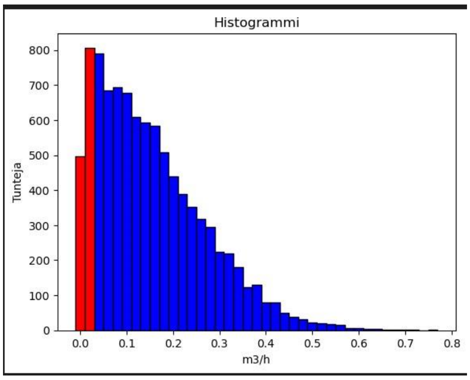
Python-code for this solution can be found from GitHub:sta.

Practical Application of Problem Solving

A clear threshold temperature was identified, above which heating buildings no longer requires circulating hot water in the radiator network. Thus, the potential savings from the annual reduction of heat losses can be estimated with reasonable accuracy. In this particular example, the housing company concluded that the annual savings potential is approximately 240 EUR/year. A device that can control water circulation based on weather forecasts appears to already be available on the market. An example of such a device is the Weather Compensation by Danfoss.

The lowest selling price of the device seems to be around 450 EUR, so without installation costs, the payback period for the device is slightly less than 2 years.

Heat Losses in Warm Circulation Water Systems and Their Potential Annual Savings



To interrupt the circulation of warm water, it is necessary to predict, with sufficient accuracy (e.g., 95% confidence), the hours when circulation is not needed. This is particularly challenging in situations where water consumption in the residential building is measured using only one meter. However, it is helpful to focus on predicting only those moments when hourly water consumption is zero or very low, for example, 10-20 liters per hour. Flushing a toilet consumes about 10 liters of water, so such consumption during nighttime is likely just the use of cold water.

The forecast was made using the FB Prophet algorithm, which achieved a prediction accuracy of 53%, a level that is far from sufficient. Since the result is actually only slightly better than a coin toss, I deemed it appropriate to run the data through a logistic regression algorithm as well. This showed a possible improvement in prediction accuracy to 59.9%, which is a modest enhancement. To achieve an even better outcome, I decided to test a third algorithm. I chose the Gradient Boosting algorithm because it has proven to be the best classifier in several Kaggle competitions. I obtained a training data accuracy of 67% with this algorithm. However, the test data accuracy remained at 61.3%. The size of the gap between these two figures indicates that the model is starting to overfit. Without overfitting, the training data accuracy could reach 63.5%.

Since the accuracy of the models remains significantly below the target value with several algorithms, it is necessary to consider other ways to improve accuracy. Feature engineering could be one alternative method. Without additional variables, either organically obtained or created through feature engineering, it is unlikely to achieve the desired result.

Therefore, no solution to this problem was found with any of the algorithms, but the Python codes for the tested algorithms can be found on GitHub.

Teachings of the project

The project was educational in many ways. During its course, there was an opportunity to gain practical experience with several different algorithms, which I understand reflects real-world scenarios. When searching for a sufficiently good algorithm to achieve a set goal, it is not necessarily reached with the first chosen algorithm. Switching from one algorithm to another ultimately isn't very difficult or labor-intensive. Examples of different algorithm applications can be found, for instance, on Kaggle. Sklearn also provides useful examples.

Another aspect that I found beneficial in the project was the practical learning of hyperparameter usage. They allow for fine-tuning the performance of the algorithm, which can further increase accuracy.

A third thing that was perhaps less covered during the course was how to determine overfitting or underfitting of the algorithm. This is also a very useful aspect to consider for the future.

The fourth thing that this project taught was that data processing, before it can be fed into the algorithm for training, takes surprisingly much time.

The fifth lesson learned was that sometimes existing metrics do not accurately reflect the desired outcome, making it worthwhile to consider better metrics to evaluate the performance of the algorithm.

Additionally, the project aimed to test running data through algorithms in the Azure environment. However, there wasn't enough time for this, as data processing and learning about the algorithms took more time than expected.