INNOVATIVE MECHANICAL ENGINEERING **ISSN 2812-9229 (Online)**
University of Niš, Faculty of Mechanical Engineering
VOL. 4, NO 1, 2025, PP. 1-10

Original scientific paper *

DAY-AHEAD FORECASTS OF EXCHANGED HEAT IN A DISTRICT HEATING SUBSTATION WITH ENSEMBLE METHODS

Mirko M. Stojiljković[1], Marko G. Ignjatović[1], Goran D. Vučković[1]

[1]University of Niš, Faculty of Mechanical Engineering in Niš, Niš, Serbia

Abstract. *Short-term forecasts of heating loads are very important for planning the operation of energy supply systems successfully. The heat demand of a building depends on various parameters that include the materials, geometry, occupancy, type of activity, etc. There are a number of machine learning methods and approaches used in the literature to accomplish these tasks. This paper presents a segment of broader research. The objective is to forecast the amount of heat exchanged in a district heating substation between the primary network and the secondary heating system of a multi-story residential building. Forecasts are performed 24 hours ahead, with the resolution of one hour. The paper applies and compares multiple ensemble regression methods based on decision trees. The input parameters are heat demand lags, time-related variables, e.g. hour of day, day of week and month, and dry bulb temperature as the most important weather variable. The time-series problem is transformed into a classical supervised machine learning problem. The models are trained and tested with the actual measured data collected over five heating seasons. The paper examines the performance of four methods: gradient boosting, histogram gradient boosting, extremely randomized trees and random forest. The applied evaluation metrics for the models are the coefficient of determination, root mean squared error and mean absolute error. All methods used have very similar prediction performance. Random forest has the smallest root mean square error (43.56 kWh), while extremely randomized trees have the lowest mean absolute error (27.34 kWh).*

Key words: *Decision tree ensembles, District heating, Forecast, Heating load*

1. INTRODUCTION

Modern district heating (DH) systems rely on heat demand forecasts for optimal operation planning, management and renewable energy utilization. Data-driven models are often suitable for this task because they have sufficient speed and accuracy. In addition,

***Received: April 02, 2025 / Accepted April 16, 2025.**
Corresponding author: Mirko M. Stojiljković
Institution: University of Niš, Faculty of Mechanical Engineering in Niš, Niš, Serbia
E-mail: mirko.stojiljkovic@masfak.ni.ac.rs

they do not require the values of all the variables that impact the demand, such as building material properties, geometry, weather data occupancy and usage patterns, etc.

One of the main characteristics of such problems is the existence of complex and non-linear dependencies of the target variable and predictors, especially weather data and user behavior. Shakeel et al. [1] underlined the importance of short-term load prediction for DH systems and used a hybrid of FB-Prophet that enables feature extraction and light gradient boosting machine (LightGBM) that is efficient in handling nonlinearities. The best approach examined for LightGBM hyperparameter optimization was the grid search. In addition, they emphasized the specific properties of DH-load-related data: the absence of consistent trend and seasonality, as well as several weather parameters as inputs.

There are several supervised machine learning models that can be applied to the problem of time series forecasting of heat demand. Frison et al. [2] stated that accurate predictions of the heat demand are important to improve energy efficiency and achieve economic benefits in DH systems. They compared several types of artificial neural networks (ANN) with the seasonal autoregressive integrated moving average with exogenous regressors (SARIMAX) and concluded that a convolutional neural network (CNN) is the most accurate one, while all models learned the trends successfully, but exhibited difficulties with fluctuations and peaks. Leiprecht et al. [3] predicted the heat demand of a DH system for a 72-hour horizon with weather forecast input and concluded that long short-term memory (LSTM) and adaptive boosting (AdaBoost) perform better than SARIMAX.

Saloux and Candanedo [4] emphasized the importance of short-term prediction of the heat load for the optimal management of renewable energy systems with energy storage. They used a decision tree (DT), support vector machine (SVM) and ANN to predict the heat demand six and 48 hours ahead. An important conclusion is that weather forecasts can improve performance to some extent for short-term prediction. They used outdoor air temperature, solar radiation, time of day, working hours and weekends as input. Similarly, Runge and Saloux [5] predicted the demand six and 24 hours in advance and concluded that the applications with input related to weather forecast performed better. LSTM and extreme gradient boosting (XGBoost) are found to be the methods with the best performance within a number of grid-searched hyperparameter combinations, while the latter is much faster to train. Morteza et al. [6] predicted electricity and heat demand with one hour resolution using deep recurrent neural network (RNN) and paid special attention to hyperparameter optimization, as a complex and time-consuming task. In this case, RNN performed better than gradient boosting (GB).

There are two types of time series forecast problems: (1) one-step ahead predictions where only the demand for the next time step is obtained and (2) multiple-step ahead or multistep predictions that output a series of future time steps. There are two strategies for multistep predictions [7, 8]: (1) the recurrent strategy that uses a series of one-step predictions where the predicted value from the previous step is used as an input for the next step and (2) the direct strategy that applies the models with multiple outputs or multiple models. Xue et al. [8] presented multistep predictions of the heat load with XGBoost, SVM and deep ANN using outdoor temperature and lag demands as inputs. They compared the direct and recursive strategies and concluded that the recursive strategy performs better. Zdravković et al. [7] demonstrated the effectiveness of the LSTM encoder-decoder architecture for single- and multistep forecasts. They applied local interpretable modelagnostic explanations (LIME) and concluded that the most important features are the heat amounts one and 23 hours ago, followed by the outdoor temperature for the previous hour.

Trabert et al. [9] predicted the heat demand and return water temperature of a DH system to optimize the operation of heat storage and cut the peaks. They found that the accuracy of the weather forecast is a very significant factor for prediction and peak shaving. Potočnik et al. [10] compared multiple predictive models for multistep forecast of the heat demand for a DH system. The forecast horizon was 48 hours and the resolution was one hour. Gaussian process regression (GPR) yielded the best results. The authors found the relevance of temperature forecasts for the model accuracy, but did not confirm the significance of future solar irradiation. They also concluded that the data for the last four days are useful for prediction.

The divide-and-conquer strategy is sometimes convenient for time series predictions of the heat demand. Kurek et al. [11] compared several methods for the hourly forecast on a 72-hour horizon and concluded that ANN with autoregressive input has the most accurate performance. They examined a DH network that covers the demand for heating and domestic hot water operating throughout the year. Therefore, the winter, intermediate and summer seasons were separated. The intermediate season was the hardest to predict. Another example of problem division is given by Hua et al. [12], who proposed an approach that combines linear regression (LR) with clustering and showed that its performance is close to ANN. The initial LR model was fitted and the observations were divided into clusters according to the residuals. A new regression model was defined for each cluster. They argued that clustering is not more demanding than tuning ANN hyperparameters and underlined the interpretable nature of the LR-based approach. Jesper et al. [13] used k-means clustering with regression, to predict an annual heat load with a daily resolution.

Chung et al. [14] proposed the approach that joins CNN and LSTM to learn spatio-temporal properties and improve the predictions of the heat demand to operate a cogeneration system. Alabi et al. [15] integrated CNN and LSTM with reinforcement learning (RL) for day-ahead predictions and energy system management. Chen et al. [16] combined heat demand predictions and deep-RL-based control to enhance the reliability of a system with a heat pump and water storage.

Bünning et al. [17] reduced the variance in the accuracy of the forecast for the day obtained with ANN using (1) autocorrelation of errors based on the predictions from previous days and (2) online learning by retraining the model with new data after each day.

Time-series forecasts of the heat demand are not limited to DH substations or systems. For example, Zhang et al. [18] used ANN to predict the heat demand with hourly resolution on a national level.

This paper applies and compares ensemble models based on DTs for day-ahead prediction of the heating load, i.e. the amount of heat exchanged in a DH substation between the primary network and the secondary heating system, related to a substation in the DH system in Niš, Serbia. The substation provides heat to a multi-story residential building. The time step is one hour. The predictions are based on easily available past data, such as the dry-bulb outdoor temperature, heating load values from the previous intervals and time-stamp data (month, day of week, hour of day, parts of day, etc.). Applied machine learning methods are GB, histogram GB (HGB), extremely randomized trees (ERT) and random forest (RF).

2. METHODOLOGY

Predicting a heating load using historical data as inputs and data-driven black-box models is a supervised machine learning problem. In particular, it is a regression problem because the target variable, i.e. the heating load, is continuous.

The load depends on many factors, such as building geometry and materials, weather conditions, occupancy, user habits, etc. Some of these factors are hard to express numerically with sufficient accuracy. The most widely used features for data-driven prediction of the heating load are the historical, i.e. previous values of the load, time-related variables and weather data, primarily air temperature and solar irradiation.

In this paper, chosen predictors should be suitable for day-ahead forecasts and relatively convenient to collect and prepare. The following input variables are considered:

- Heating load lags, i.e. historical load values,
- Time-stamp data: year, month, one-hot-encoded day of week, hour of day, as well as four binary variables that correspond to the one-hot-encoded time of day (morning 05:00–08:00, day 08:00–20:00, evening 20:00–23:00 and night 23:00–05:00),
- Weather data, which is limited only to the historical dry-bulb temperatures of the outdoor air due to the convenience of collection.

The final choice of the inputs is conducted based on the correlation and auto-correlation between the variables.

Feature engineering includes applying one-hot-encoding to the categorical variables, as already mentioned. It might also cover feature scaling, although this is usually not required for DT-based methods.

Formulating forecast models based on measured data often requires the data cleaning step because of measurement errors and missing observations. This is particularly important for time-series predictions. There are several approaches to the imputation of missing data. For example, the moving average imputation approach is based on the rolling window and inserts the mean or median of the nearby values. Exponential moving average puts higher weights on the recent data. There are various methods to handle the outliers. For example, the Hampel filter uses the median absolute deviation within nearby values to detect outliers and then replaces them with the median of these nearby values.

The entire dataset should be divided into two parts: (1) observations used for training, i.e. formulating data-driven models and (2) observations for testing, i.e. final validation of the model prediction performance. If the dataset is large enough, its training part could be further divided to perform cross-validation. In this paper, the last year is used for the validation of the model and previous observations are applied for fitting.

The models are formulated using four machine learning methods: (1) GB, (2) HGB, (3) ERT and (4) RF. All these methods use ensembles of DTs, i.e. combine multiple DTs as weak learners to obtain accurate predictors. GB and HGB are so-called boosting methods: they train the models sequentially focusing on the errors found in the previous models. ERT and RF train multiple DT models independently and aggregate the results when predicting.

Hyperparameter optimization is the process of tuning model hyperparameters, such as the number of DT estimators, allowed depth of DTs, minimal number of samples in a leaf of a DT, number of features used for each split, etc. Grid search is a widely used approach for hyperparameter tuning. It examines all possible combinations of provided hyperparameter values and selects the model with the best performance.

The metrics used to express and compare predictive performance of the obtained models used in this paper are the coefficient of determination (R^2), root mean square error (RMSE) and mean absolute error (MAE).

The coefficient of determination is dimensionless and represents the part of the target variance explained by the predictors. It is defined in Eq. (1):

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(y_t - \widehat{y_t})^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2} \tag{1}$$

where n is the sample size, i.e. the number of observations, y_t is the measured heating load during the time step t, expressed in kWh, $\widehat{y_t}$ is the predicted load for the interval t and $\bar{y}$ is the arithmetic mean of all measured values. In this case, an observation is related to a single time step.

RMSE represents the quadratic error of the prediction. In this paper, it is expressed in kWh. It is defined in Eq. (2):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \widehat{y_t})^2} \tag{2}$$

MAE is an absolute error. It is defined in Eq. (3) and also given in kWh:

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n}|y_t - \widehat{y_t}| \tag{3}$$

3. Results and discussion

The dataset for the day-ahead forecast of the heating load is obtained from an actual data collection system and is related to a DH substation located in Niš, Serbia. The substation receives heat from the DH plant via the network and supplies a multi-story residential building. The DH system operates from October to April. Usually, there is no heat supply during night, except when the outdoor temperature is very low.

The auto-correlation plot shown in Fig. 1 presents the values of the Pearson correlation coefficient, which measures the strength of linear relationship. It indicates that the most relevant historical value of the heating load for predictions is the 24-hour lag. The load information from approximately the same time during the previous day is strongly correlated to the future load with the correlation coefficient of 0.71. The 48-hour lag also has a very high correlation coefficient, but it can be assumed that it is mostly the consequence of its high correlation with the 24-hour lag. The 23 and 25-hour lags are also highly correlated with the predicted value. The 1-hour lag, i.e. the load from the previous hour is very relevant as well but cannot be used for day-ahead prediction. The Spearman correlation, which represents a measure of the monotonic relationship, has very similar values and trends of the coefficients as the Pearson correlation.

Fig. 2 illustrates the load dependence on the outdoor temperature with the regression line. The decreasing trend is expected. There is a large number of zero values, which dominantly correspond to the periods when the DH system does not work, e.g. during nights. These values might be particularly problematic to learn by prediction models. Unproportionate high values are mostly related to the morning peak demand.

The following features are selected as the input variables to prediction models:

- The heating load 24 and 25 hours before the time of the prediction,
- The outdoor air temperatures 24 and 25 hours before the time of the prediction,
- The month of the year,
- The day of week,
- The hour of day,
- The time of day.

All categorical features, except the hour of day, are modified to the groups of binary features using one-hot encoding.

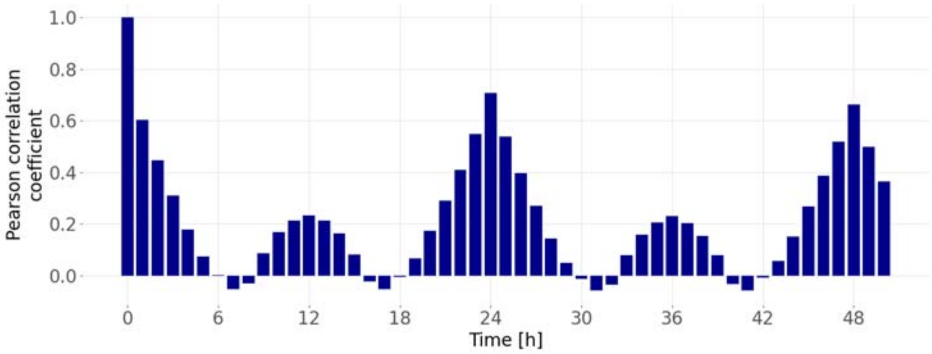

Fig. 1 Auto-correlation plot for the heating load.

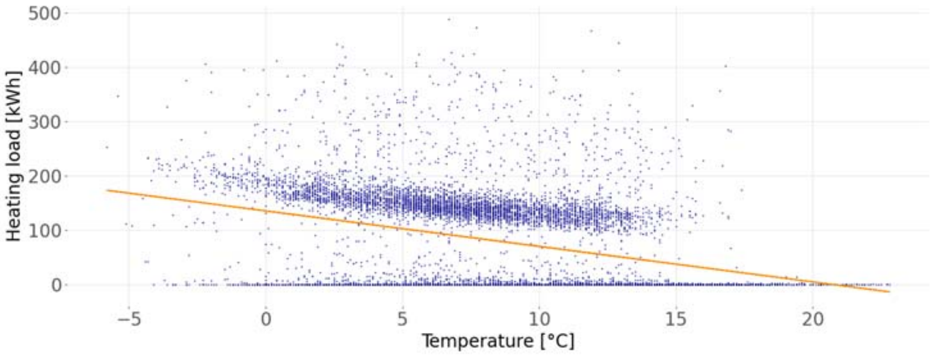

Fig. 2 Heating load dependence on the temperature.

The models are formulated and validated using four regression methods: (1) GB, (2) HGB, (3) ERT and (4) RF. Hyperparameters are optimized with the grid search approach. The predictive performance of the best model for each method is shown in Tab. 1. All methods yield very similar performance. The coefficient of determination for the validation set is almost constant across the models (approximately 0.69). RF has the lowest RMSE and the second lowest MAE. ERT performs best in terms of MAE. However, both RMSE and MAE are very similar for all four models. Therefore, the RF model is selected as the one with the best performance and the rest of the results are related to it.

Table 1 Performance of the models

Method	R^2	RMSE	MAE
Gradient boosting	0.69	43.77	28.68
Histogram gradient boosting	0.69	43.87	27.74
Extremely randomized trees	0.69	44.09	27.34
Random forest	0.69	43.56	27.50

Machine learning methods based on DTs calculate feature importance naturally. It illustrates the relative contribution of each feature to the predictive performance of the model. It is obtained by observing performance gain when a particular feature is used for data splitting. Fig. 3 shows feature importance obtained with RF. The feature with the highest importance is the heating load that occurred 24 hours before the time of prediction (56%), followed by the hour of day (14%), 25-hour load lag (5%), temperatures, etc. This means that the model learns dominantly by trying to replicate previous load patterns and much less by trying to find the relation between the heating load and the temperature.

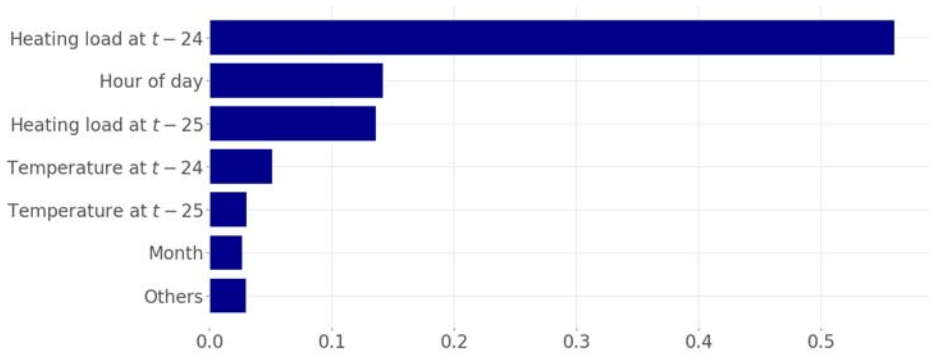

Fig. 3 Feature importance.

Fig. 4 shows the relationship between the actual and predicted heating load values, as well as the error distribution. There is a certain number of observations where the actual load is zero or very close to zero, while the predicted values are up to 200 kWh. Around 25% of all actual zero values are predicted to be over 25 kWh and around 14% have the error over 50 kWh. Handling such observations with greater attention seems to be a way to improve the predictive performance of the models. The errors are distributed around zero, approximately equally towards positive and negative values. Around 82% of errors have the absolute value lower than 50 kWh and around 57% has the error below 20 kWh.

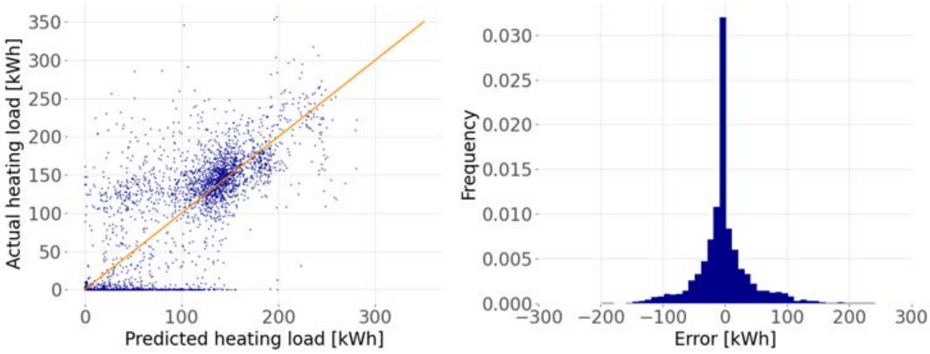

Fig. 4 Heating load forecast precision and error frequency.

Fig. 5 illustrates the distribution of errors according to the time of day. The largest errors are made in the morning hours, especially around 05:00 and 06:00, when morning peaks occur. The smallest errors correspond to the night hours when the heating is usually off. Relatively large errors occur also around noon and in the early afternoon, which is usually related to the decision of the DH operator to turn off the service due to mild conditions.

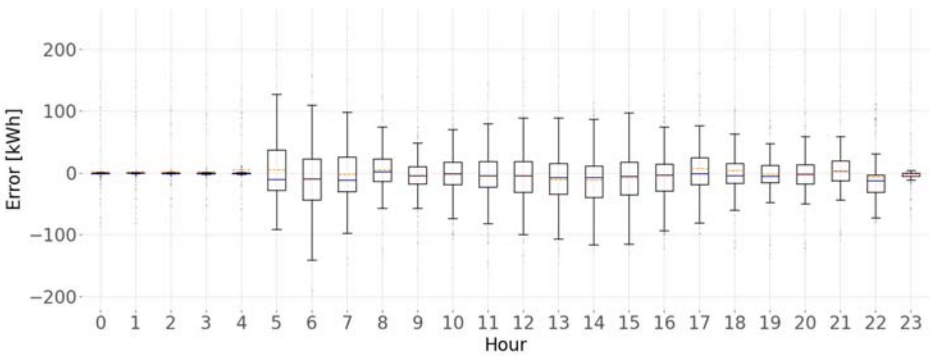

Fig. 5 Heating load forecast precision and error distribution during daytime.

Fig. 6 compares actual and predicted load profiles from November to January. It can be noticed that daily trends and changes are generally learned well, especially zero loads during nights, morning increases and evening decreases. However, a few cases with high deviations can be noted. For example, at the beginning of December and January, the load profiles differ significantly compared to the other days. In such cases, the model tries to replicate the dominant load patterns and makes large errors. The model also struggles to correctly understand how to predict the morning peak values. There are observations with both overestimated and underestimated peaks, which are consistent with the results presented in Fig. 5. Finally, it seems that the model cannot predict turning off the supply around noon.

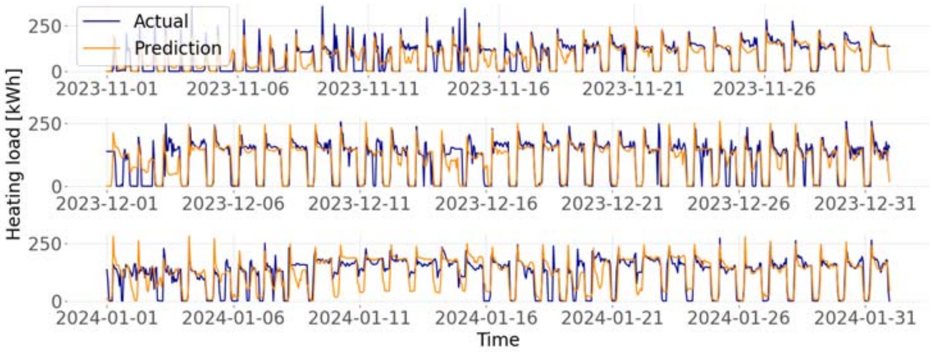

Fig. 6 Comparison of actual and predicted heating load profiles.

4. Conclusion

The forecasts of the heating load are important for optimal operation planning, management and renewable energy utilization within district heating systems. Machine learning models are often a suitable choice for this task. This paper uses and compares four ensemble regression methods based on decision trees to conduct day-ahead prediction of the amount of heat exchanged in a district heating substation between the primary network and the secondary heating system.

The main conclusions are:

- All ensemble methods exhibit very similar prediction performance and aggregation methods are slightly better than boosting methods.
- The most important features are the load values from the previous day and the hour of day.
- Dominant patterns are generally learned well.
- Large errors are often related to unusual daily load patterns, peak loads and turning off the district heating service around noon.

Future work can focus on improving accuracy by addressing the main sources of errors, applying other machine learning methods, and using stacking and voting ensembles. It can also include the estimation of the impact of prediction model performance on the operation parameters and energy savings in the district heating system.

Acknowledgement: *This research was supported by the Science Fund of the Republic of Serbia, Grant No. 23-SSF-PRISMA-206, Explainable AI-assisted operations in district heating systems - XAI4HEAT.*

References

1. Shakeel, A., Chong, D., Wang, J., 2023, *District heating load forecasting with a hybrid model based on LightGBM and FB-Prophet*, Journal of Cleaner Production, 409, 137130. https://doi.org/https://doi.org/10.1016/j.jclepro.2023.137130

2. Frison, L., Gölzhäuser, S., Bitterling, M., Kramer, W., 2024, Evaluating different artificial neural network forecasting approaches for optimizing district heating network operation, Energy, 307, 132745. https://doi.org/https://doi.org/10.1016/j.energy.2024.132745
3. Leiprecht, S., Behrens, F., Faber, T., Finkenrath, M., 2021, *A comprehensive thermal load forecasting analysis based on machine learning algorithms*, Proc. The 17th International Symposium on District Heating and Cooling, Energy Reports, 7, pp. 319–326. https://doi.org/https://doi.org/10.1016/j.egyr.2021.08.140
4. Saloux, E., Candanedo, J.A., 2018, *Forecasting district heating demand using machine learning algorithms*, Proc. 16th International Symposium on District Heating and Cooling, DHC2018, 9–12 September 2018, Hamburg, Germany, Energy Procedia, 149, pp. 59–68. https://doi.org/https://doi.org/10.1016/j.egypro.2018.08.169
5. Runge, J., Saloux, E., 2023, *A comparison of prediction and forecasting artificial intelligence models to estimate the future energy demand in a district heating system*, Energy, 269, 126661. https://doi.org/https://doi.org/10.1016/j.energy.2023.126661
6. Morteza, A., Yahyaeian, A.A., Mirzaeibonehkhater, M., Sadeghi, S., Mohaimeni, A., Taheri, S., 2023, *Deep learning hyperparameter optimization: Application to electricity and heat demand prediction for buildings*, Energy and Buildings, 289, 113036. https://doi.org/https://doi.org/10.1016/j.enbuild.2023.113036
7. Zdravković, M., Ćirić, I., Ignjatović, M., 2022, *Explainable heat demand forecasting for the novel control strategies of district heating systems*, Annual Reviews in Control, 53, pp. 405–413. https://doi.org/10.1016/j.arcontrol.2022.03.009
8. Xue, P., Jiang, Y., Zhou, Z., Chen, X., Fang, X., Liu, J., 2019, *Multi-step ahead forecasting of heat load in district heating systems using machine learning algorithms*, Energy, 188, 116085. https://doi.org/10.1016/j.energy.2019.116085
9. Trabert, U., Pag, F., Orozaliev, J., Jordan, U., Vajen, K., 2024, *Peak shaving at system level with a large district heating substation using deep learning forecasting models*, Energy, 301, 131690. https://doi.org/10.1016/j.energy.2024.131690
10. Potočnik, P., Škerl, P., Govekar, E., 2021, *Machine-learning-based multi-step heat demand forecasting in a district heating system*, Energy and Buildings, 233, 110673. https://doi.org/https://doi.org/10.1016/j.enbuild.2020.110673
11. Kurek, T., Bielecki, A., Świrski, K., Wojdan, K., Guzek, M., Bialek, J., Brzozowski, R., Serafin, R., 2021, *Heat demand forecasting algorithm for a Warsaw district heating network*, Energy, 217, 119347. https://doi.org/https://doi.org/10.1016/j.energy.2020.119347
12. Hua, P., Wang, H., Xie, Z., Lahdelma, R., 2024, *District heating load patterns and short-term forecasting for buildings and city level*, Energy, 289, 129866. https://doi.org/https://doi.org/10.1016/j.energy.2023
13. Jesper, M., Pag, F., Vajen, K., Jordan, U., 2021, *Annual industrial and commercial heat load profiles: Modeling based on k-means clustering and regression analysis*, Energy Conversion and Management: X, 10, 100085. https://doi.org/https://doi.org/10.1016/j.ecmx.2021.100085
14. Chung, W.H., Gu, Y.H., Yoo, S.J., 2022, *District heater load forecasting based on machine learning and parallel CNN-LSTM attention*, Energy, 246, 123350. https://doi.org/https://doi.org/10.1016/j.energy.2022.123350
15. Alabi, T.M., Lu, L., Yang, Z., 2024, *Real-time automatic control of multi-energy system for smart district community: A coupling ensemble prediction model and safe deep reinforcement learning*, Energy, 304, 132209. https://doi.org/https://doi.org/10.1016/j.energy.2024.132209
16. Chen, M., Xie, Z., Sun, Y., Zheng, S., 2023, *The predictive management in campus heating system based on deep reinforcement learning and probabilistic heat demands forecasting*, Applied Energy, 350, 121710. https://doi.org/https://doi.org/10.1016/j.apenergy.2023.121710
17. Bünning, F., Heer, P., Smith, R.S., Lygeros, J., 2020, *Improved day ahead heating demand forecasting by online correction methods*, Energy and Buildings, 211, 109821. https://doi.org/https://doi.org/10.1016/j.enbuild.2020.109821
18. Zhang, M., Millar, M.-A., Chen, S., Ren, Y., Yu, Z., Yu, J., 2024, *Enhancing hourly heat demand prediction through artificial neural networks: A national level case study*, Energy and AI, 15, 100315. https://doi.org/10.1016/j.egyai.2023.100315