An Empirical Analysis of Preprocessing Techniques for Short-Term Electricity Demand Forecasting

Gibak Kim[*°], Ji Eom[*], Chaehee Park[**]

ABSTRACT

This paper analyzes the impact of preprocessing techniques – including encoding, scaling, and engineered features – on the performance of short-time electricity demand forecasting based on machine learning models and validates their effectiveness through statistical hypothesis testing. We evaluated the effects of input data encoding (label, one-hot, cyclical), standardization, and engineered features. Through rigorous experiments with multiple model instances per condition, statistical significance was verified via Wilcoxon signed-rank tests. The results demonstrate that preprocessing techniques generally lead to a statistically significant improvement in electricity demand forecasting performance. However, the experimental results confirmed that the degree of effectiveness varies depending on the specific machine learning model employed. This study empirically highlights the importance of input data preprocessing in short-term electricity demand forecasting and provides insights into effective feature handling strategies considering model characteristics.

Key Words : Short-term load forecasting, Machine learning preprocessing, Statistical significance testing

I. Introduction

Accurate electricity demand forecasting is crucial for stable power system operation and planning. This study focuses on short-term forecasting, which aims to predict electricity demand for the following day. Short-term electricity demand is complexly influenced by various factors such as historical electricity consumption patterns, weather conditions, and calendar information like day of the week and holidays.

Recently, various machine learning models have been actively researched and applied in the field of electricity demand forecasting. It has been revealed that the preprocessing methods for input features significantly impact prediction accuracy. Preprocessing techniques include encoding categorical features, scaling features, and feature engineering using domain knowledge.

Prior studies have emphasized the importance of these preprocessing techniques. Researches experimentally investigated the effects of various scaling methods on classification and regression models, reporting that standardization scaling generally shows high performance[1,2]. Mahajan et al. suggested that the encoding method for cyclical features like day of the week or time can influence prediction performance depending on the type of machine learning model[3]. Zhu et al. compared and analyzed the performance of categorical feature encoders in classification and regression tasks, theoretically proving and empirically demonstrating through extensive experiments that the optimal encoder can vary depending on the model type[4]. Extensive research has also explored the use of engineered features from weather data. For instance, Kim and Lee proposed a regression model incorporating the sensitivity of electricity demand to

•° First Author and corresponding: Soongsil University School of Electrical Engineering, imkgb27@ssu.ac.kr, 종신회원
* Soongsil University Department of Physics, sally06261@naver.com
** Soongsil University School of Electrical Engineering, 7cherryhee7@naver.com

hourly temperature variations, which outperformed baseline models[5]. Spichakova et al. reported high forecasting performance using derived weather indices such as Cooling Degree Days (CDD) and Heating Degree Days (HDD), noting stronger correlations with electricity demand compared to raw temperature values [6]. Shin and Kim demonstrated the possibility of performance improvement by adding CDH (Cooling Degree Hours), perceived temperature, and discomfort index as features[7]. Time series models such as ARIMA (AutoRegressive Integrated Moving Average) have also been used in conjunction with CDD and HDD to improve demand forecast accuracy[8,9].

However, most of these studies have focused on the effects of individual preprocessing techniques or have been limited to specific models. We aim to empirically analyze the impact of major preprocessing techniques – encoding, scaling, and feature engineering – on the prediction performance in daily electricity consumption forecasting based on machine learning models. In addition, using various machine learning models such as Random Forest (RF), Artificial Neural Network (ANN), and Support Vector Regression (SVR), we will examine how the effects of preprocessing techniques appear differently depending on the model. Finally, based on the experimental results, we aim to verify whether the prediction performance difference before and after applying each preprocessing technique is statistically significant by applying statistical verification methods.

The three models (RF, SVR, ANN) were selected because they are widely used and they represent fundamentally different learning mechanisms: tree-based ensemble model, kernel-based regression, and neural network-based function approximation. This diversity allows us to assess the impact of preprocessing techniques across model families with fundamentally different characteristics. Although deep learning models like LSTM and Transformer variants have demonstrated strong performance in electricity demand forecasting, our study intentionally focuses on conventional machine learning.

This study distinguishes itself in the following key aspects compared to prior studies addressing the role of preprocessing in electricity demand forecasting[5-9]:

- Comprehensive coverage:
  We systematically explore multiple categories of preprocessing – encoding methods, scaling strategies, and weather-based feature engineering – within a single experimental framework.
- Statistical validation of performance effects:
  Unlike previous studies that report raw performance metrics, our analysis tests the statistical significance of preprocessing-induced performance differences to ensure robust and reliable conclusions.
- Model-agnostic evaluation:
  The proposed methodology assesses the impact of preprocessing across three representative model types with fundamentally different learning mechanisms, thereby increasing the generalizability of the findings.

Ⅱ. Data and Preprocessing Methods

2.1 Data Preparation

In this study, we utilized hourly electricity demand data from 2013 to 2016 provided by Shikoku Electric Power Company in Japan. Each record includes a timestamp (e.g., 2016/07/05 19:00) and the load values range from a minimum of 1.99 million kW to a maximum of 5.49 million kW, with an average of 3.555 million kW. The data exhibits strong periodic patterns, including daily and weekly cycles, which reflect typical residential and commercial electricity usage behaviors. For example, load increases during working hours and drops at night, and weekdays generally show higher demand than weekends. Seasonal effects are also present, with higher loads in winter and summer due to heating and cooling demands. The hourly electricity demand data was aggregated to obtain daily demand by summing the 24-hourly values for each day.

Given the high sensitivity of electricity demand to weather conditions, meteorological data were incorporated as important input features in the forecasting models. While electricity consumption data was aggregated over the entire island, meteorological data was collected separately for each of the four

prefectures. To integrate these, we applied a weighted average using the population ratios of each prefecture, based on the assumption that electricity usage, particularly for heating and cooling, is proportional to population size. The collected meteorological features included temperature, humidity, wind speed, and sunshine duration, all of which are known to influence electricity consumption either directly or indirectly. The hourly weather data for each city was averaged daily to generate daily meteorological features.

The training period spanned from June 1, 2013 to May 31, 2015 while the test period covered June 1, 2015 to February 28, 2016. Furthermore, as electricity demand varies seasonally, we also prepared season-specific test datasets. The summer period was defined as June through August, autumn as September through November, and winter as December through February.

2.2 Preprocessing

Electricity demand tends to vary significantly depending on calendar-related factors such as the day of the week, weekends, and public holidays. To reflect this, we added a binary feature ('is_holiday'), indicating whether each date was a holiday. The day-of-week information was represented using the 'day_of_week' feature, and the encoding method applied to this feature can affect the predictive performance of the model. We compared the following three encoding strategies:

- Label Encoding: This method maps each weekday to an integer from 0 to 6 (e.g., Monday = 0, ..., Sunday = 6). Although it appears to impose an ordinal relationship, no inherent ordering exists between weekdays, which may lead to performance degradation in some models.
- One-Hot Encoding: This approach represents each weekday as a separate binary vector. For instance, Monday is encoded as [1, 0, 0, 0, 0, 0, 0], Tuesday as [0, 1, 0, 0, 0, 0, 0], and so on. This clearly indicates the lack of an ordinal relationship between days and accurately reflects the nature of categorical data, but it has the disadvantage of increasing dimensionality.

- Cyclical Encoding: This method converts weekdays into two-dimensional coordinates using sine and cosine functions to capture their cyclical nature. For a weekday encoded as an integer, this is done using $\sin(2\pi \cdot \text{weekday}/7)$ and $\cos(2\pi \cdot \text{weekday}/7)$. This approach enables the model to naturally recognize the continuity between the end and start of a weekly cycle, which can be effective for time-related features.

When input features have different ranges, it can negatively impact model training. Therefore, we applied feature scaling to normalize the distribution of feature values. Although various methods exist for feature scaling, standardization (transforming data to have a mean of 0 and a standard deviation of 1) has been proven superior in several studies, so we adopted it as our scaling method[1,2]. Scaling can have a significant impact on distance-based algorithms such as SVR and ANN trained via gradient descent. The effectiveness of scaling under these models was empirically tested.

In addition to the basic meteorological variables such as temperature, humidity, wind speed, and sunshine duration, we generated engineered features to improve predictive performance. Engineered features can capture explicit nonlinear relationships that may not be apparent in raw variables, thereby enhancing the model's ability to learn complex patterns. In this study, the following engineered features were derived from the weather data.

- Discomfort Index (DI): The Discomfort Index is a numerical indicator that quantifies the degree of discomfort experienced by humans based on a combination of temperature (T) and relative humidity (H). Higher DI values are typically associated with increased cooling demand. The index is calculated using the following formula:

$$DI = 0.81 \times T[°C] \\ + 0.01 \times H[\%] \times (0.99 \times T[°C] - 14.3) \\ + 46.3 \tag{1}$$

- Perceived Temperature (PT): Perceived temper-

ature is calculated by combining the actual temperature (T) and wind speed (S), reflecting the extent of heat loss from the human body. This metric offers a more precise representation of the relationship between perceived thermal conditions and heating or cooling demand. The formula proposed by Environment Canada for computing perceived temperature is as follows:

$$PT = 13.12 + 0.6215 \times T[°C]$$
$$-11.37 \times \left(S[km/h]^{0.16}\right) \quad (2)$$
$$+0.3965 \times T[°C] \times S[km/h]^{0.16}$$

* Cooling Degree Hours (CDH) / Heating Degree Hours (HDH): CDH and HDH are derived by accumulating the difference between hourly temperatures (T_t) and a reference temperature (T_c), and are widely used to quantify cooling and heating loads. The formulas for CDH and HDH are as follows:

$$CDH = \sum_{t=1}^{24} \max(0, T_t - T_c)$$
$$HDH = \sum_{t=1}^{24} \max(0, T_c - T_t) \quad (3)$$

III. Experimental Design

3.1 Models and Features

The target variable for the forecasting models in this study is the daily electricity consumption for the following day. The input features are organized into three primary categories. First, we included historical electricity consumption variables: consumption one day prior (D1), three days prior (D3), and seven days prior (D7). These variables are intended to reflect both daily and weekly cycles of electricity demand, enabling the model to learn patterns such as differences between weekdays and weekends[10,11]. Second, calendar-related information was incorporated, including the day of the week and public holiday indicators. The day of the week was encoded using one of the encoding techniques described previously, while the public holiday indicator ('is_holiday') was included as a binary feature. Third, weather data and derived weather-related features were included. The raw weather variables comprised temperature, humidity,

wind speed, and sunshine duration. Additionally, engineered features such as PT, DI, CDH and HDH were added to capture non-linear and complex relationships between weather conditions and electricity demand.

To investigate how the effectiveness of preprocessing techniques varies across different machine learning models, we selected three representative regression models. The first model is RF, an ensemble learning algorithm based on decision trees. RF is capable of modeling complex non-linear relationships between features and is known for its relative insensitivity to preprocessing methods, particularly feature scaling. The second model is SVR, which employs a radial basis function (RBF) kernel to learn non-linear relationships. Due to its distance-based nature, SVR is highly sensitive to feature scaling and encoding choices. The third model is an ANN, which combines affine transformations with non-linear activation functions to learn complex patterns. The performance of ANNs is particularly influenced by the quality of feature and the scaling of input data.

RF and SVR were implemented with the scikit-learn (sklearn) package, while the ANN was implemented using Keras. For RF, we used 100 trees with no depth limitation which provides sufficient ensemble diversity while avoiding excessive computational overhead. The choice of 100 trees balances bias and variance. For SVR, the regularization parameter C=100 and epsilon=0.01 were selected. A high C enables the SVR model to fit the training data more closely, which is beneficial when short-term fluctuations carry operational significance. Likewise, a small epsilon value ensures that minor prediction errors are penalized, allowing the model to be more responsive to subtle variations in consumption patterns caused by weather or calendar effects. For the ANN, the model consists of a single hidden layer with 224 ReLU-activated neurons, trained with the Adam optimizer (learning rate=0.01, batch size=32) for 150 epochs. This shallow architecture was selected to minimize overfitting given the moderate data size, while still allowing the network to capture non-linear dependencies. Although early stopping and dropout were considered for regularization, they did not lead to noticeable improvements and were excluded from

the final configuration. We believe that the limited complexity of the ANN, consisting of only a single hidden layer, helped reduce the risk of overfitting. Accordingly, additional regularization techniques such as dropout and early stopping showed no meaningful benefit. Hyperparameters for all models were selected using grid search. The search ranges were determined based on preliminary experiments to ensure reasonable model performance across all conditions.

3.2 Evaluation Method

The predictive performance of the models was evaluated using Mean Absolute Percentage Error (MAPE). MAPE measures the absolute error between predicted and actual values, normalized by the actual value and expressed as a percentage. It offers an intuitive interpretation of the relative accuracy of the predictions.

To determine whether differences in predictive performance across preprocessing techniques were statistically significant or merely due to random variation, the Wilcoxon signed-rank test was employed. This non-parametric statistical test compares the median differences of paired samples obtained under two different conditions and does not require the assumption of normally distributed data. Consequently, it is well-suited for cases where the normality assumption may not hold or where the sample size is small. The Wilcoxon signed-rank test calculates the difference for each pair, ranks the absolute values of the differences, and then compares the signed rank sums, thus considering both the magnitude and direction of the differences. In this study, we conducted 10 repeated experiments for each experimental condition, using identical input data and model. The results of these repetitions were treated as paired samples for statistical testing of differences in MAPE between different preprocessing configurations. A significance level of p-value < 0.05 was used as the criterion for statistical significance. A p-value below 0.05 indicates that the performance difference between two methods is statistically significant.

3.3 Comparative Experiment Design

To analyze the effects of preprocessing techniques

on forecasting accuracy, the following experimental groups were designed:

- Encoding Technique Comparison: This experiment evaluates the impact of weekday encoding methods on predictive performance. The input features consist of electricity consumptions (D1, D3, D7), and weekday information. All features are standardized. The weekday information is encoded using three different methods: label encoding, one-hot encoding, and cyclical encoding. The resulting MAPE values of RF, SVR, and ANN models are compared, and statistical significance is assessed. To minimize the influence of external factors such as heating and cooling loads, weather information is excluded from this experiment, and testing is conducted using the autumn test dataset, when electricity demand is relatively less sensitive to weather fluctuations. This allows for a clearer evaluation of the encoding effects alone.

- Scaling Effect Analysis: This experiment investigates the impact of the scope of standardization on model performance. Three scaling conditions are compared: no scaling applied (no_scaling), scaling applied only to numerical features (num_scaling), and scaling applied to all features (all_scaling). The MAPE values of RF, SVR, and ANN models are compared, and statistical significance is tested. Since the effect of scaling may vary depending on the encoding method, experiments are conducted independently for cyclical encoding and one-hot encoding. As in the encoding comparison experiment, the autumn test dataset is used for analysis.

- Engineered Feature Experiment (Extended Weather Features): This experiment evaluates how adding or replacing raw weather feature with engineered weather features—DI, PT, CDH, and HDH—affects predictive performance. The input features include D1, D3, D7 electricity consumption, weekday information (encoded using cyclical encoding), public holiday indicator, and various combinations of weather-related features (basic weather

features only, with or without engineered features). This experiment is conducted using the summer and winter test datasets, where electricity demand is highly sensitive to weather factors.

Ⅳ. Experimental Results

4.1 Encoding Technique Experiments

Based on the results of the statistical significance tests, no statistically significant difference was observed between label and cyclical encoding in the RF model. However, all other pairs of encoding methods exhibited statistically significant differences in performance. Table 1 summarizes the mean predictive performance of each model for the different encoding methods.

As a tree-based model that performs optimal splits based on individual features, RF is generally less sensitive to the choice of encoding technique. Consistent with this characteristic, no significant performance difference was found between label encoding and cyclical encoding, suggesting that the model does not strongly leverage the ordinal or cyclical structure of the weekday feature. In contrast, one-hot encoding resulted in relatively lower predictive performance, and the difference was statistically significant. This may be attributed to the increased feature dimensionality introduced by one-hot encoding (seven-dimensional binary feature for weekdays), which could reduce the efficiency of the model's splitting criteria during training.

For SVR, performance differences across all three encoding methods were statistically significant, though the absolute differences in performance were not substantial. Cyclical encoding yielded the best performance, likely due to its ability to capture the inherent periodicity of the weekday feature. Similar to

Table 1. Mean MAPEs according to encoding method

Model	Label Encoding	Cyclical Encoding	One-hot Encoding
RF	2.027	2.035	2.102
SVR	2.355	2.328	2.428
ANN	2.327	2.000	1.909

RF, one-hot encoding showed the lowest performance.

In the ANN experiments, unlike the other models, one-hot encoding slightly outperformed cyclical encoding. This suggests that for neural networks, encoding methods that provide clear categorical distinctions (such as one-hot encoding) can be more effective. The ANN was likely able to leverage the expanded feature space and learn the independent binary representations of weekdays effectively within its hidden layers.

4.2 Feature Scaling Experiments

Tables 2 and 3 present the mean predictive performance (MAPE) of each model under different standardization conditions. For RF, no notable differences in performance were observed regardless of whether scaling was applied, for both encoding methods. This empirical result supports the theoretical understanding that tree-based models are inherently insensitive to the absolute scale of input features. Since decision trees split based on feature thresholds rather than magnitudes, differences in feature scales have minimal impact on performance.

In contrast, the SVR model exhibited statistically significant differences in performance across all scaling conditions for both encoding methods. The MAPE values were substantially reduced when standardization was applied (both num_scaling and all_scaling) compared to when no scaling was applied (no_scaling).

This aligns with the structural property of SVR,

Table 2. Mean MAPEs under different scaling conditions under cyclical encoding

Model	no_scaling	num_scaling	all_scaling
RF	2.035	2.035	2.035
SVR	4.086	2.383	2.328
ANN	3.611	1.926	2.000

Table 3. Mean MAPEs under different scaling conditions under one-hot encoding

Model	no_scaling	num_scaling	all_scaling
RF	2.102	2.102	2.102
SVR	4.576	2.264	2.430
ANN	3.797	1.908	1.915

which relies on distance-based calculations, making the model highly sensitive to the scale of input features. Notably, in the case of one-hot encoding, num_scaling yielded slightly better performance than all_scaling, while the opposite trend was observed with cyclical encoding, where all_scaling slightly outperformed num_scaling. This suggests that binary features from one-hot encoding, which already have values constrained to 0 or 1, may not benefit significantly from additional scaling.

For the ANN model, when one-hot encoding was used, there was no statistically significant performance difference between num_scaling and all_scaling. However, all other pairs of experimental conditions showed significant differences. As ANN models are trained using gradient descent, discrepancies in feature scales can lead to unstable learning dynamics, such as disproportionately large weight updates or slower convergence. Standardization helps mitigate these issues by ensuring consistent feature ranges, contributing to more stable training and improved convergence toward optimal solutions.

4.3 Engineered Feature Experiments

This section presents the results of experiments analyzing the impact of incorporating engineered weather features—namely, DI, PT, CDH, and HDH—in addition to the basic weather information. To evaluate the effects of each engineered feature, we compared three feature configurations: models using all basic weather variables (weather_all), models using only the engineered feature with its source variables excluded (xxx_replace), and models where the engineered feature was added to the basic weather variables (xxx_plus). "xxx" denotes the specific engineered feature under consideration.

Discomfort index (DI), derived from temperature and humidity as shown in Equation (1), is expected to correlate positively with cooling demand; thus, performance improvements were anticipated for the summer dataset. The results are presented in Table 4.

In the RF experiments, no statistically significant differences were observed between DI_replace and DI_plus in both summer and winter datasets.

Table 4. Mean MAPEs for different DI configurations

Model		weather_all	DI_replace	DI_plus
RF	Summer	2.268	2.220	2.225
	Winter	2.854	2.813	2.802
SVR	Summer	2.896	2.832	2.848
	Winter	3.309	3.241	3.242
ANN	Summer	2.281	2.214	1.915
	Winter	2.073	2.071	1.945

However, the inclusion of discomfort index led to a slight but statistically significant improvement in performance compared to using basic weather variables alone.

In the SVR experiments, no significant differences were found between DI_replace and DI_plus for the winter dataset. In contrast, for the summer dataset, DI_replace significantly outperformed DI_plus. This result aligns with the earlier observation that SVR is highly sensitive to the number and composition of input features. Using a compact, information-rich feature like discomfort index alone appears to better support SVR's stability and generalization, by reducing the risk of overfitting in high-dimensional kernel space.

In the ANN experiments with winter data, no statistically significant difference was observed between weather_all and DI_replace. However, using the DI consistently resulted in significant performance improvements overall. The configuration with DI_plus achieved the lowest MAPE, suggesting that the neural network was able to effectively leverage the additional feature through its hidden layer representation learning.

Perceived temperature (PT), derived from temperature and wind speed as shown in Equation (2), was evaluated as summarized in Table 5. For RF, adding perceived temperature (PT_plus) produced better results than replacing basic features with it (PT_replace), with a particularly notable improvement over weather_all in the winter dataset. This finding is consistent with the fact that perceived temperature reflects heat loss due to wind and thus aligns more closely with actual power consumption patterns during heating-dominant winter periods.

Table 5. Mean MAPEs for different PT configurations

Model		weather_all	PT_replace	PT_plus
RF	Summer	2.268	2.282	2.260
	Winter	2.854	2.830	2.799
SVR	Summer	2.896	2.789	2.856
	Winter	3.309	3.005	3.326
ANN	Summer	2.281	2.174	1.952
	Winter	2.073	2.094	1.937

In SVR experiments, statistically significant differences were observed for all pairs of experimental conditions. As with the discomfort index results, replacing basic weather features with perceived temperature (PT_replace) yielded better performance than adding it (PT_plus), again highlighting SVR's sensitivity to feature composition.

Similarly, in ANN experiments, significant differences were observed across all experimental pairs. Unlike SVR, the ANN benefited more from adding perceived temperature to the existing feature set (PT_plus), resulting in improved predictive performance.

CDH and HDH depend on the choice of temperature and time window. While these metrics can be calculated over all 24 hours, it is also possible to restrict HDH calculation to working hours (9:00 – 18:00), which may better capture winter power consumption patterns, as residential heating often involves non-electric sources. Based on preliminary experiments, this study used 24°C as the reference for CDH and 11°C for HDH. CDH was calculated over the full 24-hour period, while HDH was computed using working-hour temperatures only.

This reflects the observation that in winter, workplace electricity demand (rather than residential) tends to dominate. CDH experiments used summer data; HDH experiments used winter data. The results are presented in Tables 6 and 7.

Across all models, adding CDH or HDH generally resulted in the best predictive performance. For RF, the addition of CDH (CDH_plus) did not yield statistically significant improvements over using basic weather variables alone. Similarly, no significant difference was observed between HDH_replace and

Table 6. Mean MAPEs for different CDH configurations

Model	weather_all	CDH_replace	CDH_plus
RF	2.268	2.447	2.251
SVR	2.896	2.961	2.732
ANN	2.281	2.344	1.682

Table 7. Mean MAPEs for different HDH configurations

Model	weather_all	HDH_replace	HDH_plus
RF	2.854	2.826	2.759
SVR	3.309	3.435	3.229
ANN	2.073	2.392	1.973

HDH_plus.

Ⅴ. Conclusions

This study confirms that preprocessing decisions significantly influence model performance for short-term electricity demand forecasting. For instance, SVR and ANN achieved significantly lower MAPE values with cyclical encoding or one-hot encoding compared to label encoding. The superior performance of cyclical encoding over one-hot encoding may stem from its ability to preserve ordinal and periodic relationships in time-related variables. This effect was especially pronounced in models like SVR, where high-dimensional sparse inputs (e.g., one-hot vectors) can degrade generalization performance. Notably, ANN exhibited the best performance with one-hot encoding, suggesting that neural networks are particularly effective when categorical variables are represented as clearly separated inputs. In contrast, RF showed minimal sensitivity to encoding methods due to its tree-based architecture.

In the scaling experiments, both SVR and ANN were highly sensitive to feature scaling, achieving the lowest MAPE when standardization was applied to all features, including categorical variables. Standardization scaling appeared to benefit ANN models by stabilizing gradient-based optimization and improving convergence, consistent with prior findings in neural network training dynamics. RF, consistent with its design, exhibited no significant performance differences based on scaling.

In the analysis of engineered features, incorporating domain knowledge-based features such as DI, PT, CDH, and HDH was shown to improve forecasting performance. ANN achieved the best accuracy when all basic weather variables and engineered features were used together, while SVR performed better when only carefully selected key engineered features were included.

This study provides practical guidelines for developing effective preprocessing strategies in real-world electricity demand forecasting applications. The systematic comparative analysis of various preprocessing configurations offers an empirical foundation for improving the accuracy of electricity demand forecasts.

While our study included limited tests of preprocessing combinations, further exploration of preprocessing interactions should be pursued in future work to uncover potentially synergistic effects. Future work will also explore how these preprocessing strategies interact with deep architectures such as LSTM or Transformer variants.

References

[1] M. Ahsan, et al., "Effect of data scaling methods on machine learning algorithms and model performance," *Technol.*, vol. 9, no. 3, pp. 52-68, 2021. (https://doi.org/10.3390/technologies9030052)

[2] L. de Amorim, G. Cavalcanti, and R. Cruz, "The choice of scaling technique matters for classification performance," *Applied Soft Comput.*, vol. 133, pp. 1-19, 2023. (https://doi.org/10.1016/j.asoc.2022.109924)

[3] T. Mahajan, G. Singh, and G. Bruns, "An experimental assessment of treatments for cyclical data," in *Proc. Computer Sci. Conf. CSU Undergraduates*, 2021.

[4] W. Zhu, R. Qiu, and Y. Fu, "Comparative study on the performance of categorical variable encoders in classification and regression tasks," *arXiv preprint arXiv: 2401.09682v1*, 2024. (https://doi.org/10.48550/arXiv.2401.09682)

[5] J. Kim and C.-Y. Lee, "The forecasting power energy demand by applying time dependent sensitivity between temperature and power consumption," *J. Soc. Korea Ind. Syst. Eng.*, vol. 42, no. 1, pp. 129-136, 2019. (https://doi.org/10.11627/jkise.2019.42.1.129)

[6] M. Spichakova, J. Belikov, and K. Nou, "Feature engineering for short-term forecast of energy consumption," in *Proc. IEEE PES: ISGT-Europe*, Bucharest Romania, 2019. (https://doi.org/10.1109/ISGTEurope.2019.8905 698)

[7] D.-H. Shin and C.-B. Kim, "A study on deep learning input pattern for summer power demand prediction," *J. KITT*, vol. 14, no. 11, pp. 127-134, 2016. (https://doi.org/10.14801/jkiit.2016.14.11.127)

[8] S.-W. Jung and S. Kim, "Electricity demand forecasting for daily peak load with seasonality and temperature effects," *Korean J. Applied Statistics*, vol. 24, no. 5, pp. 843-853, 2014. (https://doi.org/10.5351/KJAS.2014.27.5.843)

[9] C. Choo, H. Joo, and E. Hwang, "Time series analysis for prediction intervals of electricity demand with weather variables," *J. Korean Data Analysis Soc.*, vol. 26, no. 2, pp. 471-488, 2024. (https://doi.org/10.37727/jkdas.2024.26.2.471)

[10] J.-Y. Oh, et al., "Short-term load forecasting using XGBoost and the analysis of hyperparameters," *Trans. KIEE*, vol. 68, no. 9, pp. 1073-1078, 2019. (https://doi.org/10.5370/KIEE.2019.68.9.1073)

[11] J. Jeon, "Deep learning based short term load forecasting considering holidays and special days," M.S. Thesis, Soonchunhyang University, 2022.

Gibak Kim

1994 : B.S. degree, Seoul National University
2007 : Ph.D. degree, Seoul National University
2011~Current : Professor, School of Electrical Engineering, Soongsil University
<Research Interests> Demand forecasting, Artificial intelligence, Signal processing
[ORCID:0000-0001-5114-4117]

Chaehee Park

2021~Current : Undergraduate student, Soongsil University
<Research Interests> Demand forecasting
[ORCID:0009-0008-6393-5707]

Ji Eom

2019~Current : Undergraduate student, Soongsil University
<Research Interests> Demand forecasting
[ORCID:0009-0007-7374-5184]