**Power Consumption Prediction AI Model Development**

**1. Overview**

Accurate power consumption prediction is essential for stable and efficient energy supply. This project aims to develop an AI algorithm capable of predicting power consumption by leveraging simulation models.

Objective: To develop a model that predicts power consumption at specific time points using building and spatiotemporal information.

**2. Data Information**

Training Data: Data from 100 buildings spanning from June 1, 2022, to August 24, 2022. Includes information on temperature, precipitation, wind speed, humidity, solar radiation, solar energy, and power consumption (kWh) at hourly intervals.

Building Information: Details of 100 buildings, including building number, type, total area, cooling area, solar capacity, ESS storage capacity, and PCS capacity.

Validation Data: Forecast data from August 25, 2022, to August 31, 2022, including temperature, precipitation, wind speed, and humidity.

**3. Data Preprocessing**

**(1) Initial Data Preparation**

Datetime Conversion: Convert the datetime column into a proper date-time format. Sorting and Index Reset: Ensure data consistency by grouping by building\_type, sorting chronologically, and resetting indices.

**(2) Adding Weekday and Weekend Indicators**

Weekday Indicator: Add a column weekday to represent the day of the week (0: Monday ~ 6: Sunday). Weekend Indicator: Add a column is\_weekend to indicate whether the day is a weekend (True for weekends, False for weekdays).

(3) Setting Holiday Conditions by Building Type

- Discount Mart: It was observed that discount marts have holidays that differ from regular weekends. To incorporate this, a new variable holiday was created: Holiday Rule: Discount marts observe holidays on the 2nd and 4th Sundays of each month. Week Calculation: Derived a variable week\_of\_month from the date to identify the week of the month.

Holiday Assignment: Sundays matching the above condition were assigned holiday = 1.

- Data Center: Data centers are operational 24/7, so all holiday values were set to 0.

- Other Building Types: Compared the average power consumption (power\_consumption\_kWh) on weekends and weekdays for each building type. If the average weekend power consumption was lower than weekday power consumption, weekends were assigned holiday = 1. Otherwise, weekends and all other holidays were assigned holiday = 0.

**Table 1) Number of Holidays per Building**

|  |  |  |
| --- | --- | --- |
| **building\_type** | **holiday** | |
| **0** | **1** |
| Apartment | 16320 | 0 |
| Commercial | 11712 | 4608 |
| Data Center | 10200 | 0 |
| Department Store & Outlet | 16320 | 0 |
| Discount Mart | 15360 | 960 |
| Hospital | 11712 | 4608 |
| Hotel & Resort | 16320 | 0 |
| Knowledge Industry Center | 11712 | 4608 |
| Other Building | 30600 | 0 |
| Public | 11712 | 4608 |
| Research Center | 11712 | 4608 |
| University | 11712 | 4608 |

**3. Model Application**

**(1) Data Preprocessing and Splitting**

Using the training data, we separated the features and target variable necessary for predicting power consumption.

* Feature Variables (X): A set of variables expected to influence power consumption, including temperature, wind speed, humidity, total area, cooling area, parking, month, time of day, day of the week, and whether it is a holiday.
* Target Variable (y): power\_consumption\_kWh, representing power consumption for each time period.

The training data was split into training (80%) and validation (20%) datasets to evaluate the model's ability to predict new data. This process was performed using the train\_test\_split function, with a random\_state set to ensure reproducibility.

**(2) Model Training: Random Forest and XGBoost**

Both models were used to train a power consumption prediction model. Using the training data, the models were trained via the fit method to learn rules for predicting power consumption across time periods.

**(3) Model Evaluation**

Model performance was evaluated by using the validation dataset (X\_val) to calculate predicted values (y\_pred) and comparing them with actual values (y\_val). The evaluation metrics used were Mean Squared Error (MSE), which quantifies prediction accuracy as the average squared difference between predicted and actual values, and the R-squared (R²) score.

**Table 2) Evaluation results**

|  |  |  |
| --- | --- | --- |
| Model | MSE | R^2 |
| Random Forest | 30290.97 | 0.995062 |
| XGBosst | 150683 | 0.975434 |

**(4) Model Prediction**

The selected Random Forest model was applied to the test data, and the results were output. Predicted values after the red dotted line represent those generated by the model.

**Picture 2) power consumption for building type – prediction results**

텍스트, 스크린샷, 도표, 그래프이(가) 표시된 사진

자동 생성된 설명텍스트, 라인, 도표, 그래프이(가) 표시된 사진

자동 생성된 설명

텍스트, 폰트, 스크린샷, 친필이(가) 표시된 사진

자동 생성된 설명

텍스트, 친필, 폰트, 그래프이(가) 표시된 사진

자동 생성된 설명

텍스트, 라인, 폰트, 스크린샷이(가) 표시된 사진

자동 생성된 설명

텍스트, 폰트, 스크린샷, 도표이(가) 표시된 사진

자동 생성된 설명

**5. Conclusion**

The Random Forest model accurately predicted power consumption across most building types, effectively reflecting key factors such as seasonality and time of day. However, for certain outlier data points (e.g., holidays and specific building types), additional feature engineering or model improvement is needed. These results are expected to provide practical support for power usage management and energy optimization strategies.

**Patterns in Training and Test Data:**

Power consumption patterns varied significantly by building type, with sharp fluctuations observed during specific time periods or days of the week. Additionally, identifying factors like holidays and periodicity for each type proved critical for constructing an accurate model.