1)Data preprocessing

- Data integration: The 5 different datasets are integrated.

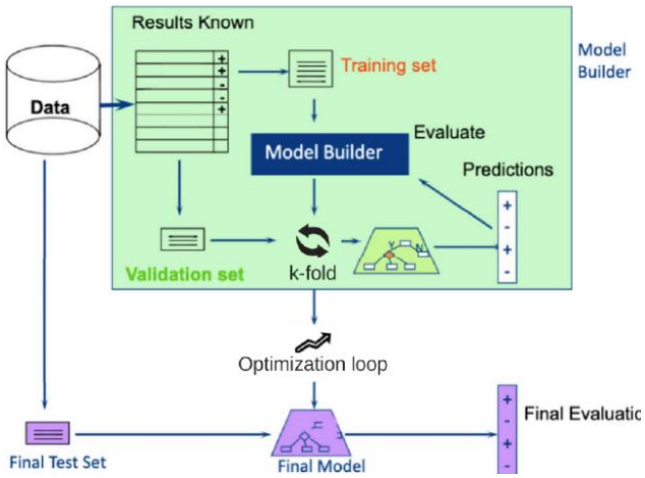
- Data cleaning: Remove missing, discarding zero reading consumptions, Outliers, Transformation (Categorization): converting categorical features to numerical

- Feature Engineering:

- Extract New Features: Max/Avg/Min of daily consumption (and any other features such as season, day, month, etc.)
- Adding Next-day Avg/Min/Max Temperature
- Adding Y class (next_day_max_consumption)

- Visualization: Students need to show some figures for showing the relation between energy consumption and other features. For instance in the solution you see that by increasing the temperature the energy consumption reduces (Scatter plot for the relation between temperature and date)

- Experimental Setup Diagram:



2) Experimental Setup

- splitting the dataset (important features? feature selection)

- Train the models

- Evaluation

Performance measures: Several performance measures can be chosen: Such as Error (MSE, RSME, MAE, and MAPE) and Accuracy.

For recommendation: Based on the fair comparison, they need to compare the models and based on the performance of them draw a conclusion that which model is better and why. For instance “Based on the experiment results, Model X is has the most accurate daily maximum energy load prediction for a 30 minutes interval than the model Y so ..”

Bonus: Compare the results of the students with the last month (December) to know how they could predict well. I provided the last month data in bonus.csv