Air Pollution Measurements Prediction

Yu Li

Nanjing University of Science and Technology

Deakin University

Chinese Academy of Sciences

(None)

Overview

Problem Definition

Data Description

Feature Engineering

Model Training

Result

Problem Definition

Air Pollution Measurements Prediction

Data Description

Train Data Description

Test Data Description

Feature Engineering

Model Training

Result

Problem Definition

Air Pollution Measurements Prediction

Data Description

Feature Engineering

Model Training

Result

Problem Definition

Air Pollution Measurements Prediction

Problem Definition

Air Pollution Measurements Prediction

Data Description

Feature Engineering

Model Training

Result

In this competition you are predicting the values of air pollution measurements over time, based on basic weather information (temperature and humidity) and the input values of 5 sensors. The three target values to you to predict are:

- target_carbon_monoxide
- target_benzene
- target_nitrogen_oxides

Problem Definition

Data Description

Train Data Description

Test Data Description

Feature Engineering

Model Training

Result

Data Description

Train Data Description

Problem Definition

Data Description

Train Data Description

Test Data Description

Feature Engineering

Model Training

Result

Elements	Number
datetime	7111
degC	408
relativehumidity	762
absolutehumidity	5451
sensor1	3882
sensor2	3882
sensor3	3882
sensor4	3882
sensor5	3882
targetcarbonmonoxide	95
targetbenzene	405
targetnitrogenoxides	3268

Test Data Description

Problem Definition
Data Description
Train Data Description
Test Data Description
Feature Engineering
Model Training
Result

Elements	Number
datetime	2247
degC	280
relativehumidity	653
absolutehumidity	1915
*sensor*1	1758
*sensor*2	1816
*sensor*3	1833
*sensor*4	1877
*sensor*5	2017

Data Visualization

Problem Definition

Data Description

Train Data Description

Test Data Description

Feature Engineering

Model Training

Result

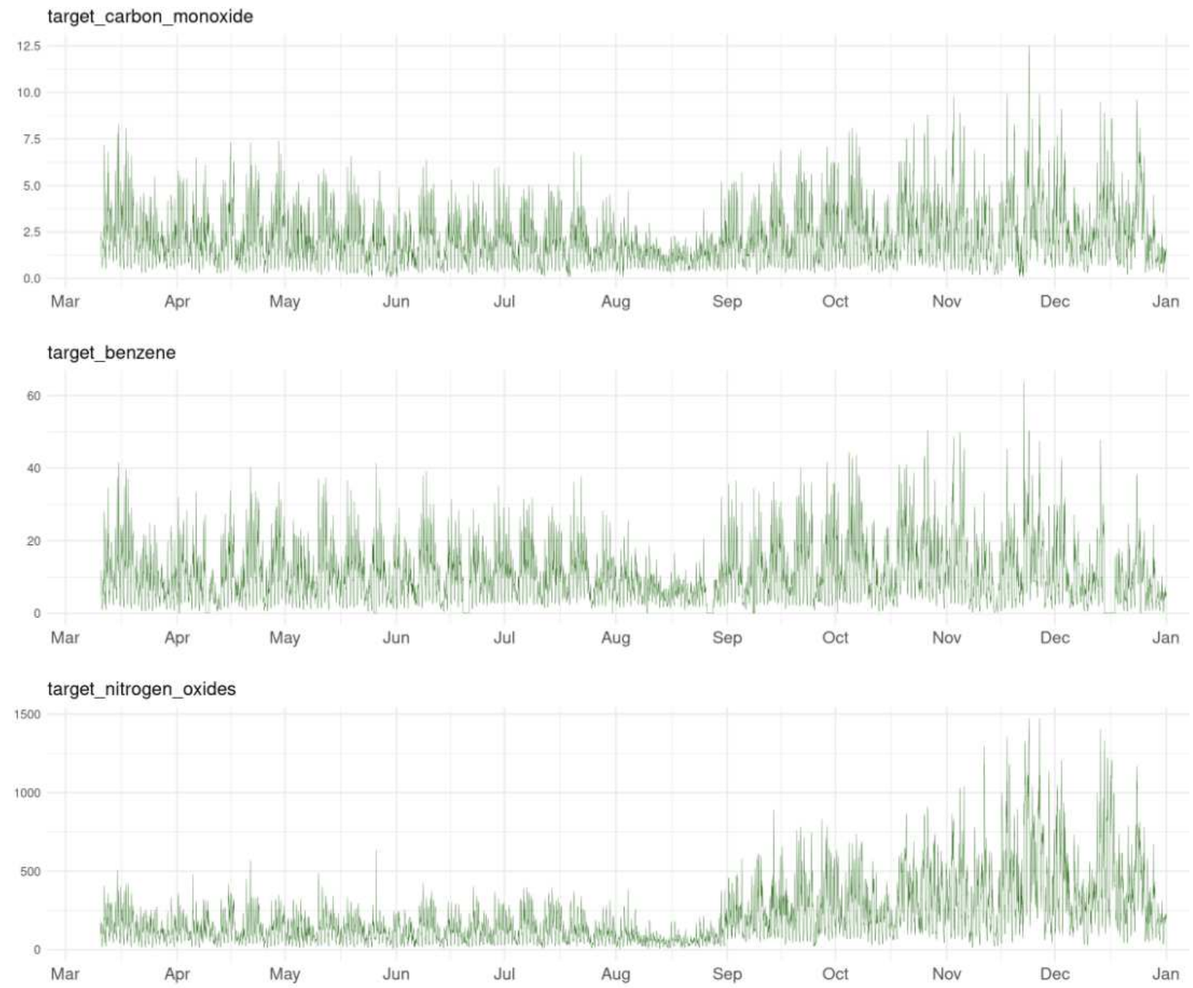

Figure 1: Target Overall Situation

Data Visualization(2)

Problem Definition

Data Description

Train Data Description

Test Data Description

Feature Engineering

Model Training

Result

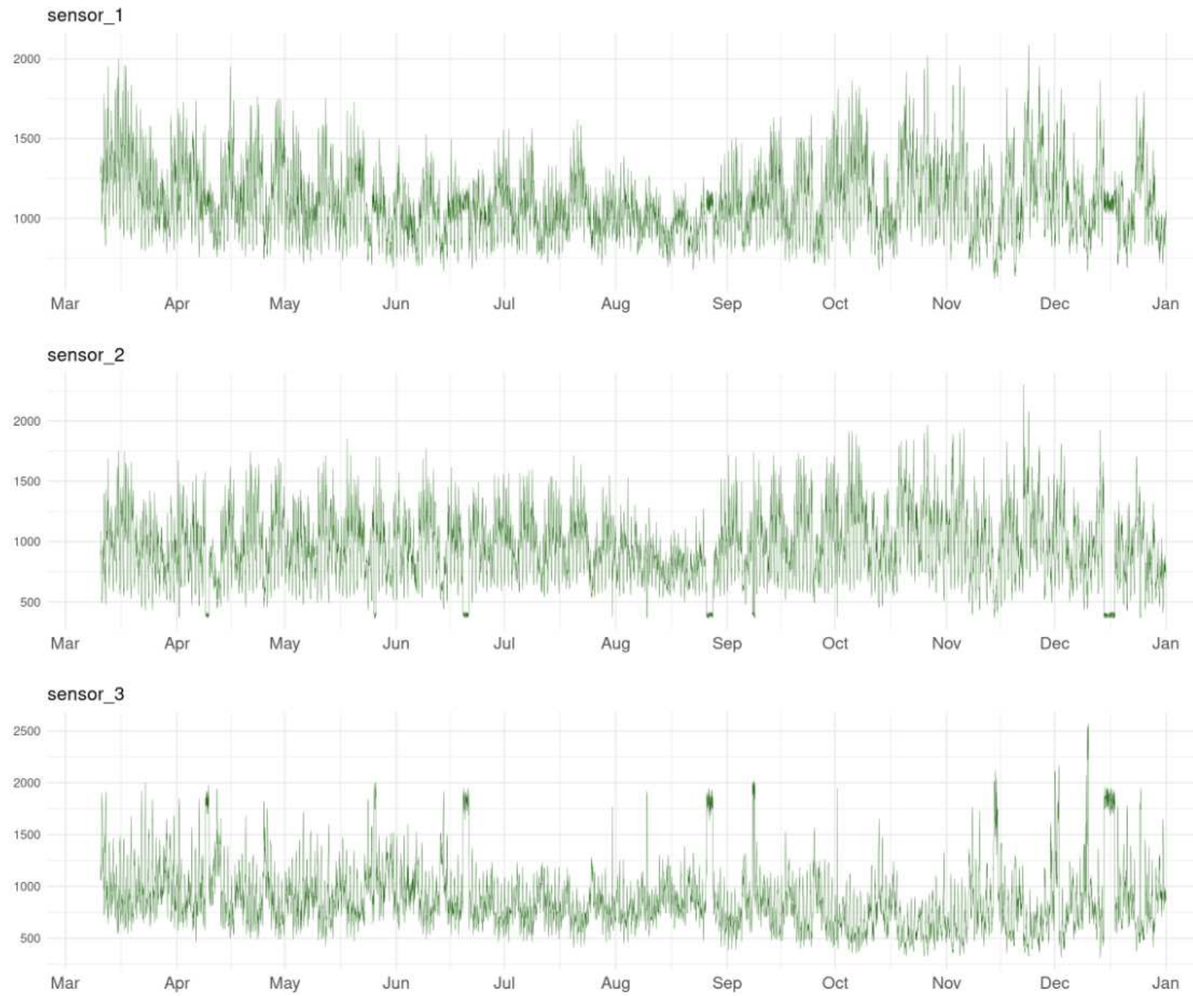

Figure 2: Sensor(1-3) Overall Situation

Data Visualization(3)

Problem Definition

Data Description

Train Data Description

Test Data Description

Feature Engineering

Model Training

Result

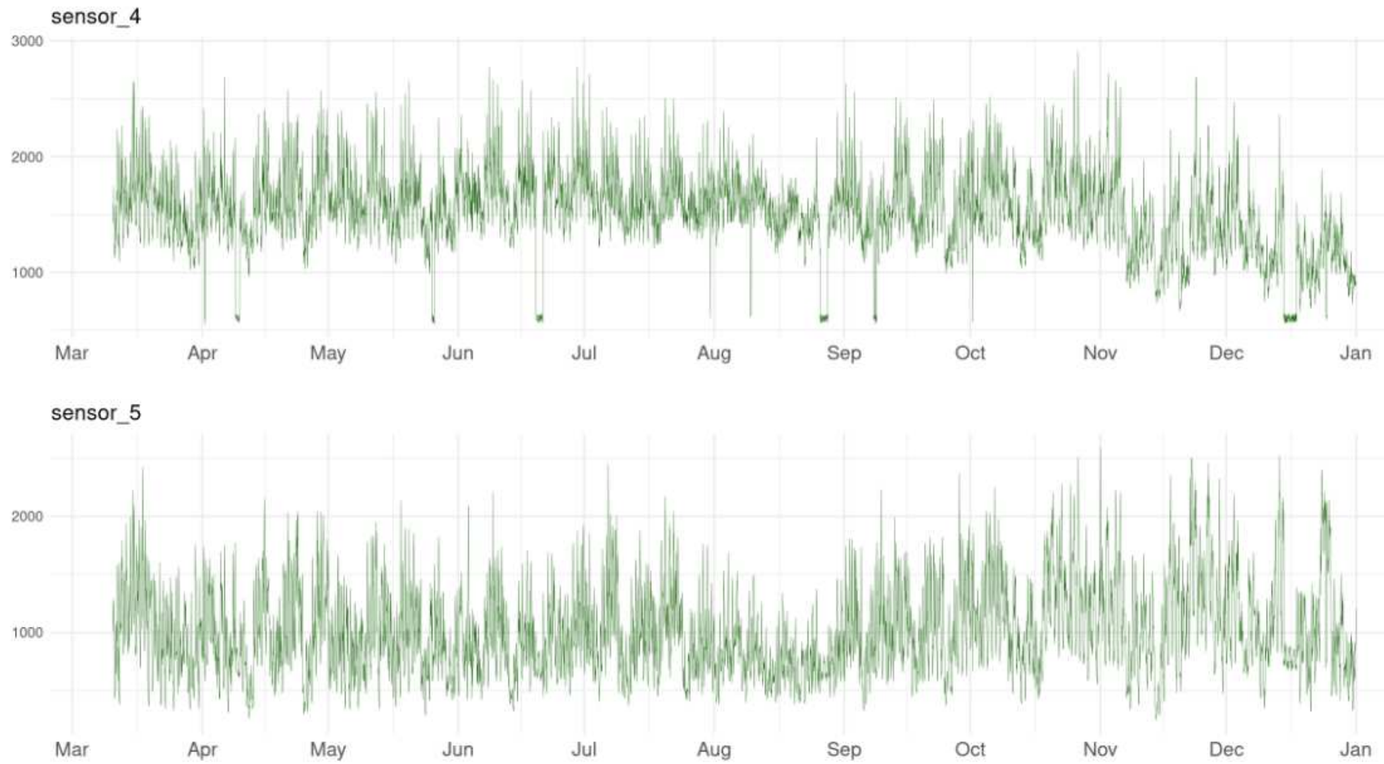

Figure 3: Sensor(4-5) Overall Situation

Data Visualization(4)

Problem Definition

Data Description

Train Data Description

Test Data Description

Feature Engineering

Model Training

Result

Figure 4: Weather Overall Situation

Data Visualization(5)

Problem Definition

Data Description

Train Data Description

Test Data Description

Feature Engineering

Model Training

Result

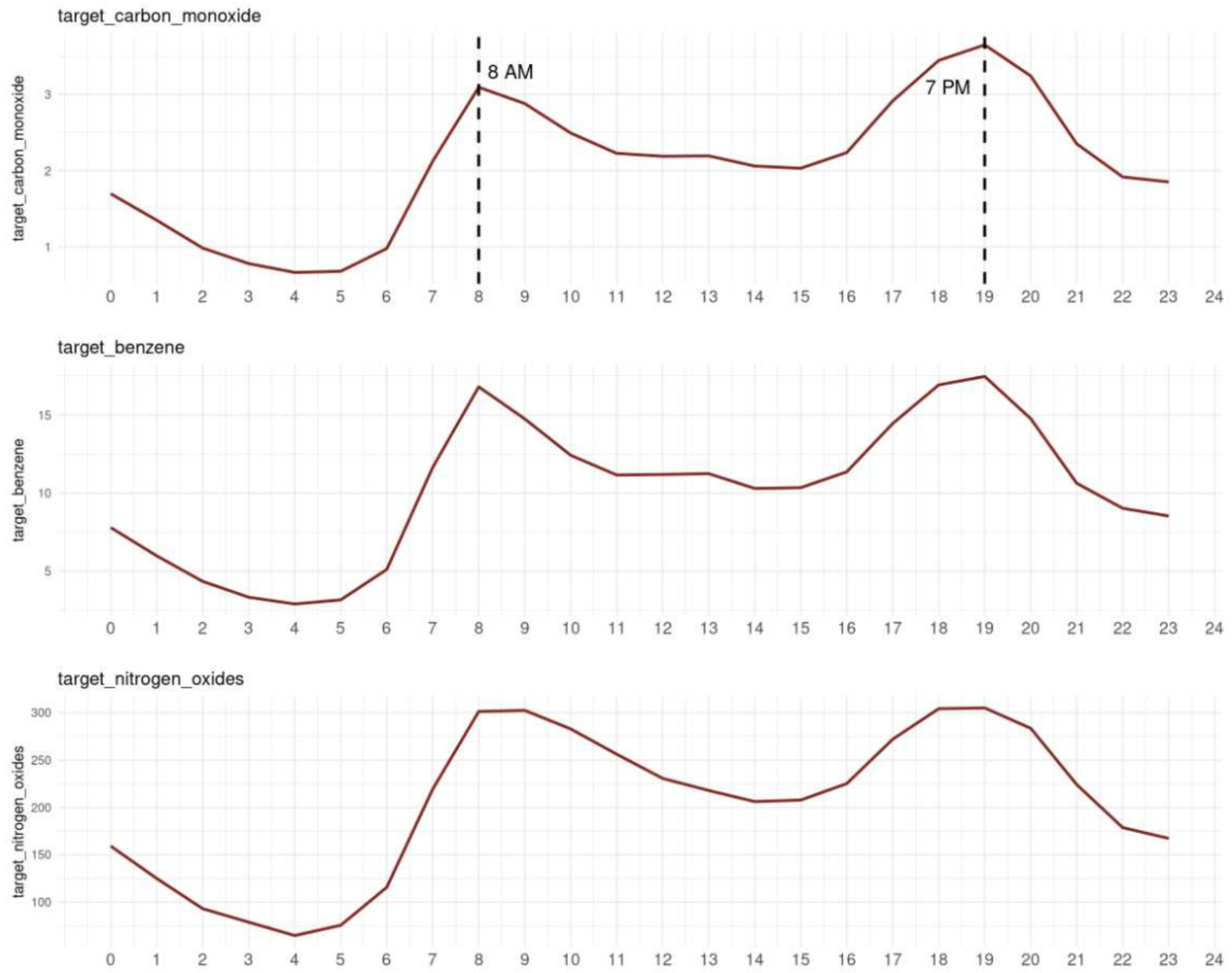

Figure 5: Target Daily Hourly Change

Data Visualization(6)

Problem Definition

Data Description

Train Data Description

Test Data Description

Feature Engineering

Model Training

Result

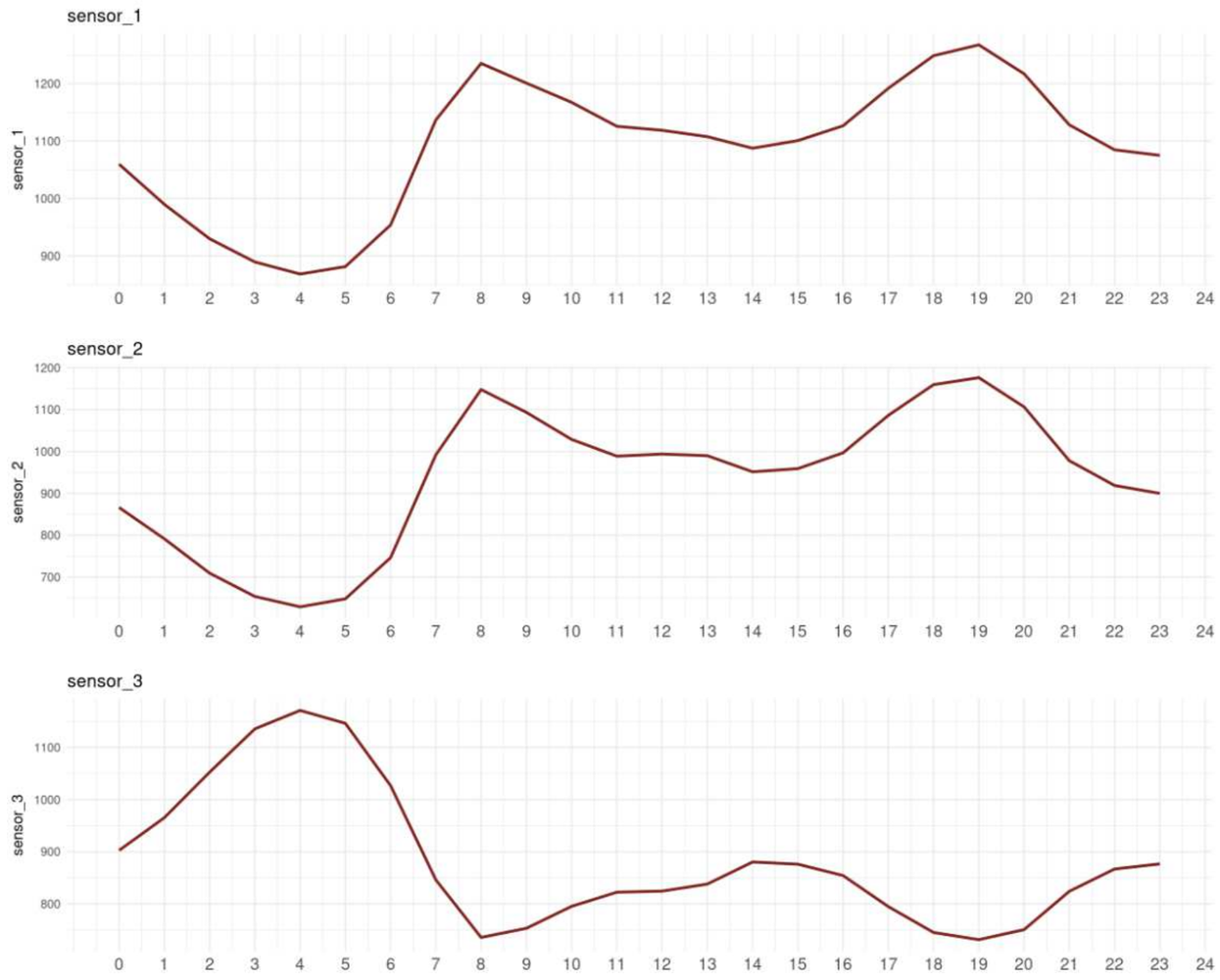

Figure 6: Sensor(1-3) Daily Hourly Change

Data Visualization(7)

Problem Definition

Data Description

Train Data Description

Test Data Description

Feature Engineering

Model Training

Result

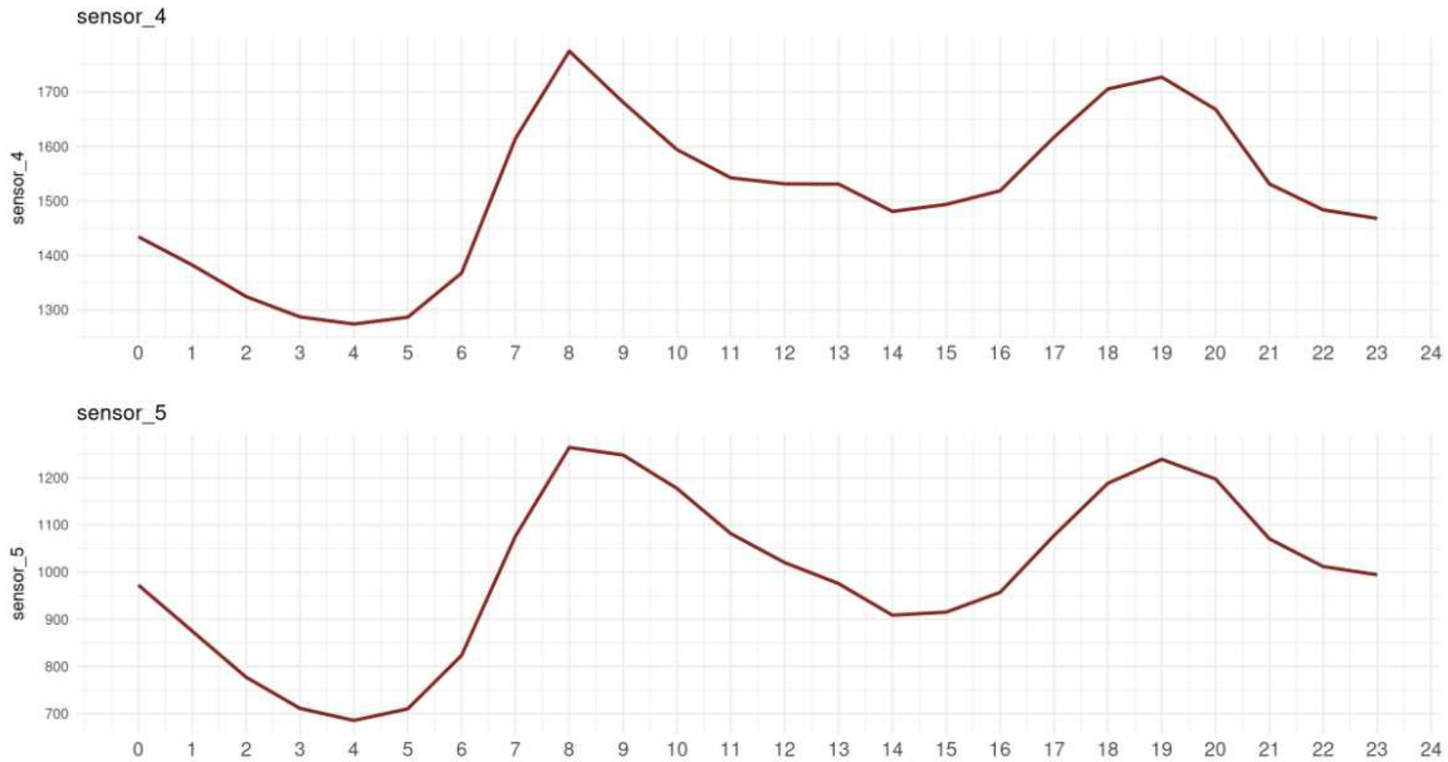

Figure 7: Sensor(4-5) Daily Hourly Change

Data Visualization(8)

Problem Definition

Data Description

Train Data Description

Test Data Description

Feature Engineering

Model Training

Result

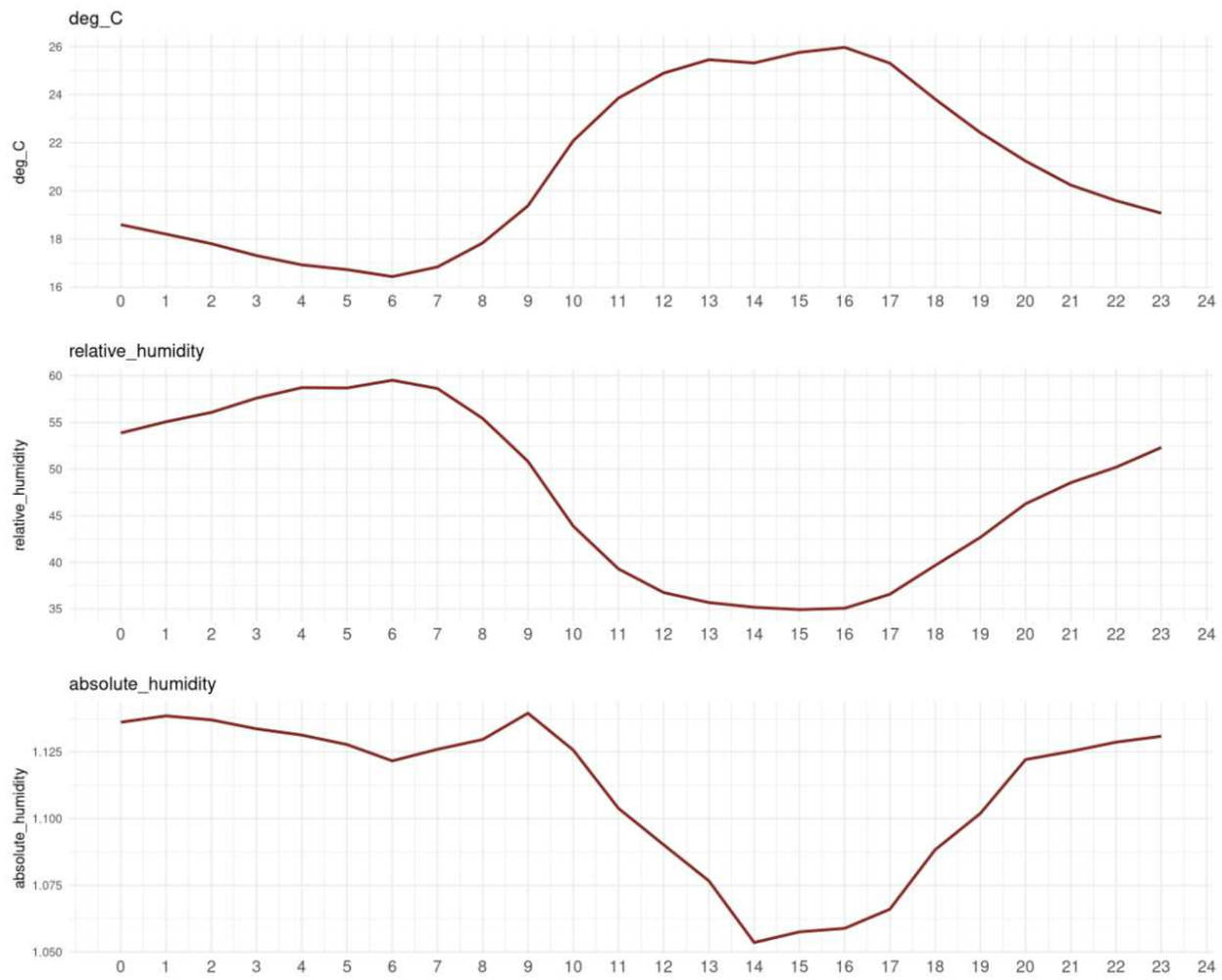

Figure 8: Weather Daily Hourly Change

Data Visualization(9)

Problem Definition

Data Description

Train Data Description

Test Data Description

Feature Engineering

Model Training

Result

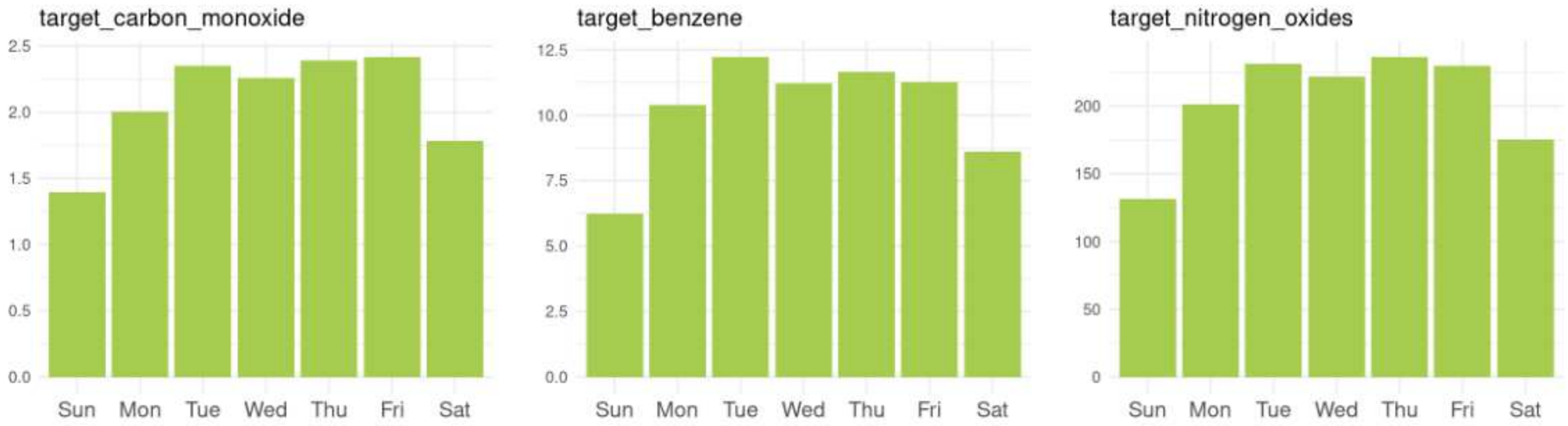

Figure 9: Target Weekly Situation

Data Visualization(10)

Problem Definition

Data Description

Train Data Description

Test Data Description

Feature Engineering

Model Training

Result

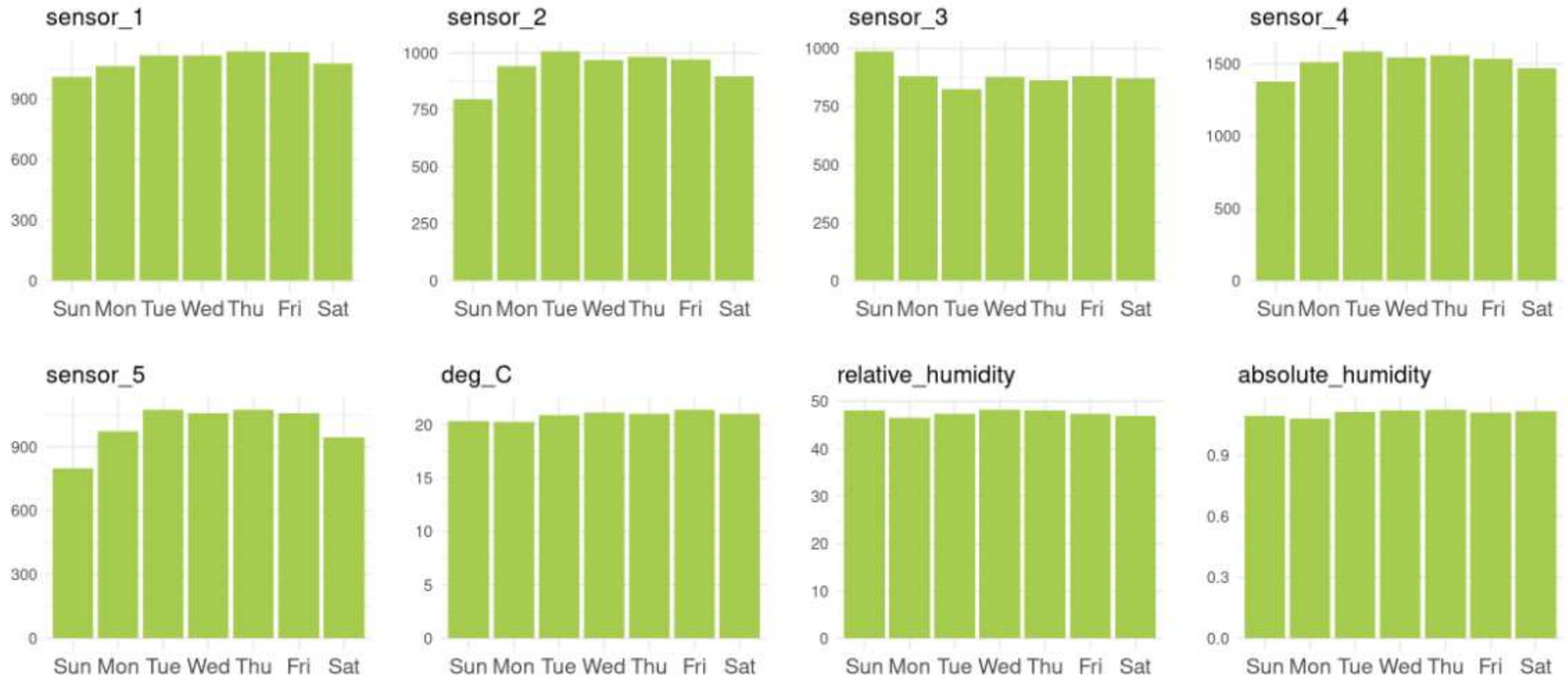

Figure 10: Sensor and Weather Weekly Situation

Problem Definition

Data Description

Feature Engineering

Model Training

Result

Feature Engineering

Feature Selection

Problem Definition

Data Description

Feature Engineering

Model Training

Result

According to the analysis of training data, the following features are used for model training:

- absolute_humidity
- deg_C
- relative_humidity
- sensor1-5
- month
- week
- is_weekend
- hour

Problem Definition

Data Description

Feature Engineering

Model Training

Result

Model Training

Model Selection

Problem Definition

Data Description

Feature Engineering

Model Training

Result

Data fitting using LGBMRegressor, the algorithm is easy to use. It only needs to put the set features and three prediction targets into the model for training, but there is no parameter optimization, which has a certain impact on the training results.

Problem Definition

Data Description

Feature Engineering

Model Training

Result

Result

Result

Problem Definition

Data Description

Feature Engineering

Model Training

Result

- Use RMSLE(Root Mean Squared Logarithmic Error) to evaluate the results.

$$RMSLE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(log(\hat{y}_i + 1) - log(y_i + 1))^2}$$

- Private Score:0.33979
- Public Score:0.387

Questions?

Problem Definition

Data Description

Feature Engineering

Model Training

Result

Contact Information

Yu Li

School of Economics and Management

Nanjing University of Science and Technology, China

2805242929@QQ.COM

TEAM FOR UNIVERSAL LEARNING AND INTELLIGENT PROCESSING