~~Title of This Paper~~

AIR POLLUTION MEASUREMENTS PREDICTION

~~AUTHOR 1~~YU LI

Abstract. 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.

Contents

Date: (None).
2020 *Mathematics Subject Classification.* Artificial Intelligence.
Key words and phrases. Machine Learning, Pollution Prediction.

1. Introduction

At a high level, what is the problem area you are working in and why is it important? It is important to set the larger context here. Why is the problem of interest and importance to the larger community?

This paragraph narrows down the topic area of the paper. In the first paragraph you have established general context and importance. Here you establish specific context and background.

"In this paper, we show that ...". This is the key paragraph in the intro - you summarize, in one paragraph, what are the main contributions of your paper given the context you have established in paragraphs 1 and 2. What is the general approach taken? Why are the specific resultssignificant? This paragraph must be really good.

You should think about how to structure these one or two paragraph summaries of what your paper is all about. If there are two or three main results, then you might consider itemizing them with bullets or in test. 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:

- e.g., First ... target-carbon-monoxide
- e.g., Second ... target-benzene
- e.g., Third ... target-nitrogen-oxides

If the results fall broadly into two categories, you can bring out that distinction here. For example, "Our results are both theoretical and applied in nature. (two sentences follow, one each on theory and application)"

Keep this at a high level , you can refer to a future section where specific details and differences will be given. But it is important for the reader to know at a high level, what is new about this work compared to other work in the area.

"The remainder of this paper is structured as follows. .." Give the reader a roadmap for the rest of the paper.Avoid redundant phrasing, "In Section 2, In section 3, ... In Section 4, ... " etc.

Test citation [**?**]. and [**?**] or **?**].

This is for , and this is for .

Number: . , , , and

We have , , the range: . 1/2.

For , as shown below:

$$a = b \times \sqrt{ab}$$

2. Preliminaries

2. Data Description

Before model training, data needs to be analyzed to determine the required features. Here is the statistics of training data and test data:

3. Method

Experiment and Analysis

~~Precision Comparison on Event Detection Methods~~

TABLE 1. Train Data Description

Elements	~~OR Event Detection~~ `Number`
datetime	~~AC Event Detection~~ 7111
degC	~~TC Event Detection~~ 408
relative − humidity	762
absolute − humidity	5451
*sensor*1	3882
*sensor*2	3882
*sensor*3	3882
*sensor*4	3882
*sensor*5	3882
target − carbon − monoxide	95
target − benzene	405
target − nitrogen − oxides	3268

TABLE 2. Test Data Description

Elements	`Number`
~~precision~~ *datetime*	~~0.83~~ 2247
degC	~~0.69~~ 280
relative − humidity	~~0.46~~ 653
~~recall~~ *absolute − humidity*	~~0.68~~ 1915
*sensor*1	~~0.48~~ 1758
*sensor*2	~~0.36~~ 1816
~~F-score~~ *sensor*3	~~0.747~~ 1833
*sensor*4	~~0.57~~ 1877
*sensor*5	~~0.4~~ 2017

In order to understand the change trend of data, the data is visualized and analyzed based on the visualization results.

3. ~~Conclusions~~

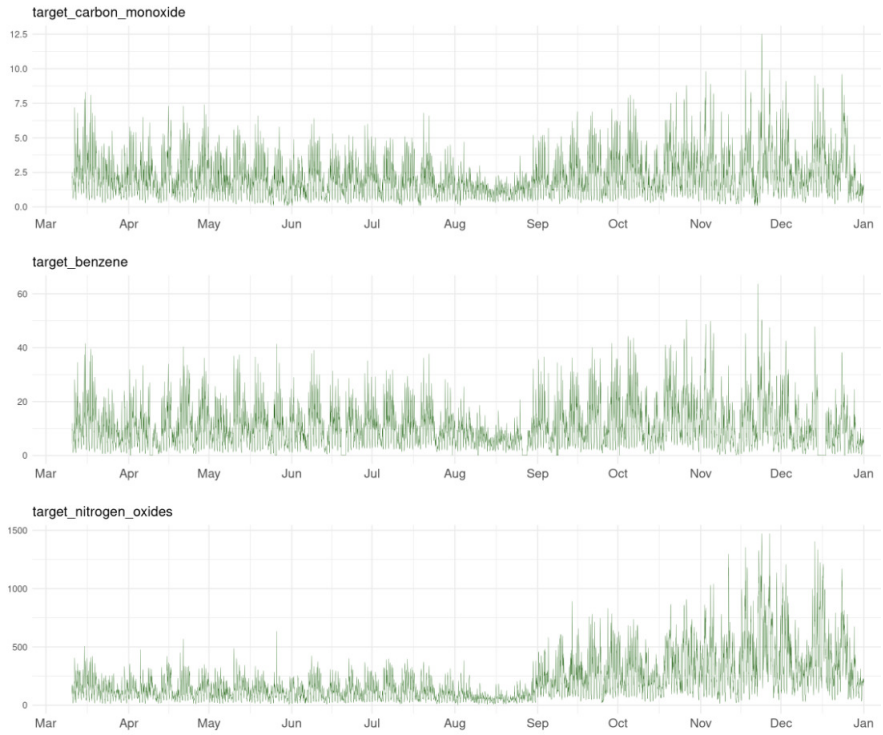

FIGURE 1. Target Overall Situation

It can be seen from the figure1 that the values of the three target pollutants in August each year will be lower, gradually rising from September, and significantly higher than the level before August, so it is necessary to take the month as a feature of the model.

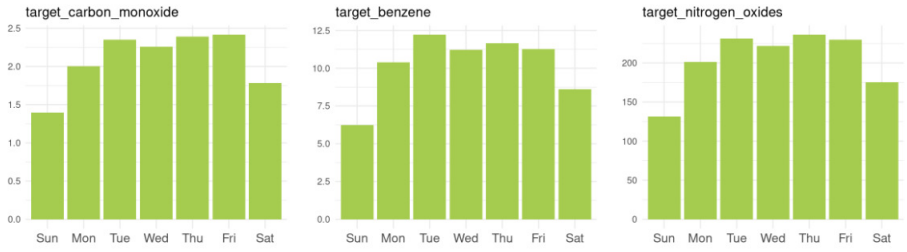

FIGURE 2. Target Weekly Situation

It can be seen from the figure2 that the content level of each pollutant at the weekend of each week will decrease significantly, so it is necessary to take whether this day is a weekend as a feature of the model.

~~Acknowledgement~~

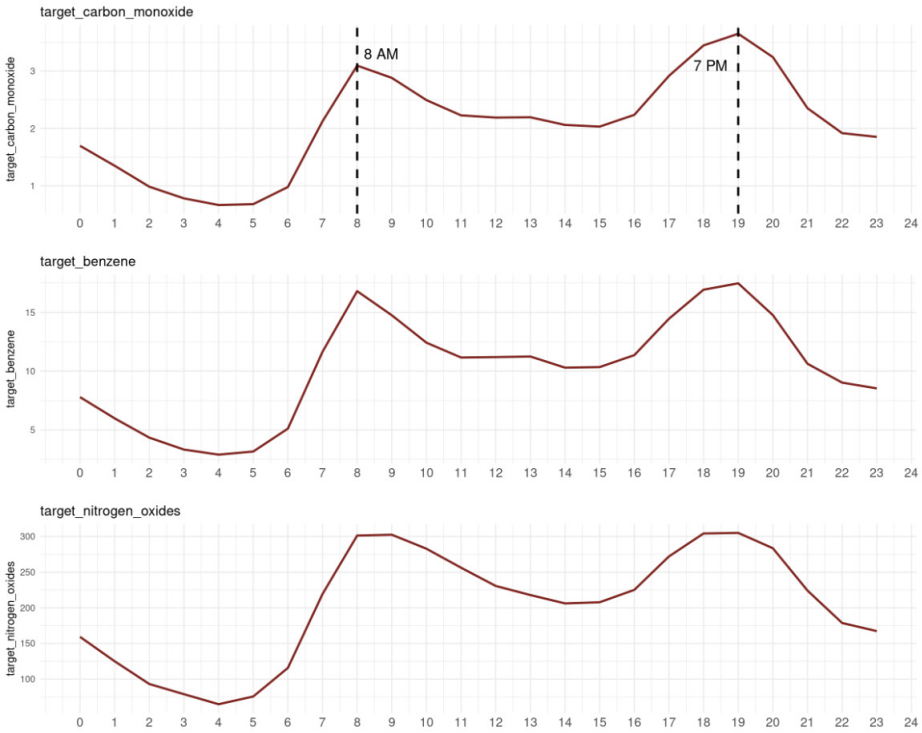

FIGURE 3. Target Daily Hourly Change

It can be seen from the figure3 that the level of each pollutant is the lowest at 5:00 a.m. every day, and then gradually rises to 8:00 a.m. to reach the first peak, and then gradually falls to 4:00 p.m., and then rises to 7:00 p.m. to reach the second peak, and then continues to decline, so it is necessary to take time as a feature of the model.

3. Feature Engineering

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

~~The authors would like to thank . . .~~

4. Model Training

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.

5. 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

(A. 1) School of Computer Science,, Nanjing University of Science and Technology, Jiangsu 246000, China

Email address, A. 1: yli@tulip.academy