1. 004: Yanyan Deng, Jiayu Zhuo, Shengyuan Gao
2. Dataset Name: Pm 2.5 data of five Chinese cities (picking up one or two cities)
3. [https://archive.ics.uci.edu/ml/datasets/PM2.5+Data+of+Five+Chinese+Cities#](https://archive.ics.uci.edu/ml/datasets/PM2.5+Data+of+Five+Chinese+Cities)
4. Source: UCI datamining Archive
5. Number of records in the dataset: 52874 (5 cities)
6. No of features: 17 (excluding the 1 No field)
7. Brief Description of the dataset:

PM 2.5 refers to atmospheric particulate matter that have a diameter less than 2.5 micrometers. The dataset contains 52874 records from five Chinese cities.

1. Table:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Attribute | Data Type | description |
| 1 | No | continuous | row number |
| 2 | year | continuous | year of data in this row |
| 3 | month | continuous | month of data in this row |
| 4 | day | continuous | day of data inthis row |
| 5 | hour | continuous | hour of data in this row |
| 6 | season | continuous | reason of data in this row |
| 7 | PM\_Dongsi | categorical | pm2.5 concentration |
| 8 | PM\_Dongsihuan | categorical | pm2.5 concentration |
| 9 | PM\_Nongzhanguan | categorical | pm2.5 concentration |
| 10 | PM\_US post | categorical | pm2.5 concentration |
| 11 | DEWP | continuous | dew point |
| 12 | HUMI | continuous | humidity |
| 13 | PRES | continuous | presure |
| 14 | TEMP | continuous | temperature |
| 15 | cdwd | categorical | combined wind direction |
| 16 | lws | continuous | cumulated wind speed(m/s) |
| 17 | precipitation | continuous | Hourly precipitation(mm) |
| 18 | lprec | continuous | Cumulated precipitation(mm) |

1. Data mining task: regression
2. What ML/DM techniques do you plan to apply?
   1. MLR or ANN
3. What do you hope to learn?
   1. We hope to find out attributes that caused pm 2.5 outbreak by using these data.
4. How can the data mining results be made use of?

Analyzing the rule of pm2.5 and reducing the frequency of air pollution(pm2.5) .