Investigation on PM2.5 Issue for Main Cities in China

NYU TANDON SCHOOL OF ENGINEERING

Dataset explanation:

Target variable: PM 2.5 density;

Time series: Jan 1st, 2010 to Dec, 31st, 2015; Data is recorded by hours;

Meteorological feature:

DEWP: Dew Point (Celsius Degree)

TEMP: Temperature (Celsius Degree)

HUMI: Humidity (%)

PRES: Pressure (hPa)

Iws: Cumulated wind speed (m/s)

Precipitation: Hourly precipitation (mm)

Problem:

Future PM2.5 density prediction

Modeling using meteorological features

	year	month	day	hour	season	PM_US Post	DEWP	HUMI	PRES	TEMP	Iws	precipitation
21	2010	1	1	21	4	NaN	-17.0	38.0	1018.0	-5.0	1.79	0.0
22	2010	1	1	22	4	NaN	-17.0	38.0	1018.0	-5.0	2.68	0.0
23	2010	1	1	23	4	129.0	-17.0	41.0	1020.0	-5.0	0.89	0.0
24	2010	1	2	0	4	148.0	-16.0	38.0	1020.0	-4.0	1.79	0.0

NYU

TANDON SCHOOL OF ENGINEERING

PM2.5 Density	Label	Value for filling
[0,100)	1	50
(100,200]	2	150
(200,300]	3	250
(300,400]	4	350
(400,500]	5	450
(500,600]	6	550
(600,700]	7	650
(700,800]	8	750
(800,900]	9	850
Over 900	10	950

What we have done:

Machine learning based modeling for filling the missing values

1. Split dataset
2. Build classification model
3. Model selection

Future work:

1. Time series analysis
2. Data summary and hypothesis

GitHub Link: https://github.com/KiyoshiKAWASAKI/FDS_project_17Fall