

1 Task Description

The health department of Lechi has noticed that pollution indicators have drastically increased. The regional government, concerned about this situation, has decided to implement a prediction model to predict the concentration on carbon monoxide in the air from the data gathered by the air-quality monitors.

In this assignment, you will use regression techniques in order to predict the CO concentration based on temporal, climatic and environmental data.

Your prediction will be evaluated according to the Mean Absolute Error (MAE) metric.

2 Dataset Description

The dataset contains hourly information of weather and environmental conditions in a time period from 2014 to 2018.

The target attribute is the carbon monoxide concentration (`carb_monox`)

Number of instances: 14000

Number of attributes: 12

2.1 Target Class:

`carb_monox`: CO concentration (micrograms per cubic meter)

2.2 Attribute Information:

Item	Attribute	Type	Values
1	hr : hour of the day	Numerical/Categorical	0-24
2	small_part : fine particulate matter concentration	Numerical	micrograms per cubic meter
3	med_part : particulate matter concentration	Numerical	micrograms per cubic meter
4	sulf_diox : Sulfur dioxide concentration	Numerical	micrograms per cubic meter
5	nitr_diox : Nitrogen Dioxide concentration	Numerical	micrograms per cubic meter
6	trioxygen : Ozone concentration	Numerical	micrograms per cubic meter
7	temp : Temperature	Numerical	degree Celsius
8	pres : Pressure	Numerical	hectopascals
9	rain : Precipitation	Numerical	millimeters
10	wind : Wind direction	Categorical	E, ENE,ESE,N,NE,NNE, NNW,NW,S,SE,SSE,SSW, SW,W,WNW,WSW
11	wind_sp : Wind speed	Numerical	meters per second
12	date : Date	Categorical	day-month-year

3 Prediction File Submission

3.1 Model Submission: 16.12.2019, 11:59am

You are asked to kindly submit the following supporting information in **two files**:

1. A brief description of the step by step methodology (i.e. pre-processing, visualization, training, testing, etc.) that you have followed to do the assignment, with the aim of illustrating the motivation behind your selected approach.

- File Format: .pdf
- Filename: **6-digit** student code (e.g. 123456.pdf)

2. The python code that you used to do the assignment, with comments within the code to ensure that they can be clearly understood.

- File Format: .ipynb, .py
- Filename: **6-digit** student code (e.g. 123456.ipynb or 123456.py)

3.2 Test Publication: 16.12.2019, 14:00pm

3.3 Prediction Submission: 18.12.2019, 23:59pm

You are kindly requested to strictly follow the described submission guidelines:

- File Format: .csv
- Filename: **6-digit** student code (e.g. 123456.csv)
- Column Format: 1 Column named **"target"**
- Row Format: Your predictions with the **same number of rows and in the same order as the test set** provided to you

4 Important:

- Check your uploaded files.
- The model can not be modified after submission.
- Any submission that does not respect the guidelines (out of deadline, empty file, wrong student code) will not be graded.
- Check your uploaded files!.