1   Task Description

In the region of Lechi, recent improvements on mental health of the population have increased the use of the public library.

In order to face this new scenario, regional government has decided to reorganize the staff and the activities carried out in the library. The library director believes that library occupancy is influenced by the weather conditions, type of day (weekend, workday) and the library staff.

In this assignment, you will use regression techniques in order to predict the use of the library by using climatic, temporal and employees information.

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

2   Dataset Description

The dataset contains an hourly information of weather conditions, workday type and library staff information. The target attribute is the library occupancy.

Number of instances: 10427

Number of attributes: 14

2.1   Target Class:

Occupancy: number of users in the library

2.2   Attribute Information:

Item	Attribute	Type	Values
1	**dteday**	Date	Date
2	**season**	Categorical	spring, summer, fall, winter
3	**hr:** hour of the day	Numerical/Categorical	0-24
4	**holiday**	Binary	0:no holiday, 1: otherwise
5	**weekday:** day if the week	Categorical	0-6
6	**workingday**	Binary	0: no working day, 1: otherwise
7	**weather**	Categorical	Clear, Cloudy, LightRain, Snow
8	**temperature**	Numerical	Degrees
9	**stemp:** temperature perception	Numerical	Degrees
10	**humidity**	Numerical	Percentage
11	**winds:** windspeed	Numerical	velocity
12	**nemployee:** average number of employees	Numerical	number
13	**expemployee:** average year of experience of employees	Numerical	years
14	**occupancy:** TARGET	Numerical	number of users in the library

3 Prediction File Submission

3.1 Model Submission: 04.12.2019, 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: 04.12.2019, 14:00pm

3.3 Prediction Submission: 06.12.2019, 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:

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