**DATA MINING PROJECT :**

**Goal :**

To forecast the air quality in future (hourly) in Beijing and London :

* the concentration of PM2.5, PM10 and O3 for 35 stations in Beijing.
* predict the concentration of PM2.5 and PM10 for 13 stations in London.

**Technics to try (It is mandatory to use three or more diﬀerent algorithms in addition to a baseline):**

<https://fr.wikipedia.org/wiki/Machine_%C3%A0_vecteurs_de_support#SVM_pour_la_r%C3%A9gression>

<https://en.wikipedia.org/wiki/Recurrent_neural_network>

<https://en.wikipedia.org/wiki/Gradient_boosting#Gradient_tree_boosting>

(techniques mentionnées intéressantes par le prof pour ce projet)

🡪discuss the diﬀerent performances you have with your methods and explain why these work or not.

**Model comparison and selection (04) :**

Using cross-validation to select the best model based on statistical signiﬁcant test

**Error Estimation :**

* Cross-validation (+ simple et + stylé à mon avis qui compte pour que dalle)
* Bootstrap