MACHINE LEARNING / APRENDIZAGEM COMPUTACIONAL — 2025/2026

MINI STUDENT'S COMPETITION ON REGRESSION DATASET

Summary

- Mini-competition: the students are asked to propose their best regression solution on a dataset provided by the Professor

Competition Assignment (individual)

The original dataset is based on FuelConsumption.csv. The categorical variables were removed, and the data were shuffled. Three new datasets were created:

- Train_FuelConsumption.csv (80% of data)
- Test_ FuelConsumption.csv (15% of data)
- Test_Simul_FuelConsumption.csv (5% of data)

You will have access only to Train_FuelConsumption.csv and Test_Simul_FuelConsumption.csv. You may use the data however you want. Test_Simul_FuelConsumption.csv is provided only as a sanity-check test (to verify that your model runs correctly). I will use Test_FuelConsumption.csv (not provided to students) to evaluate your submitted model.

Submssion

You must submit three files:

1- **Solution.py**, where:
- you present your best solution (all the steps),
- train your model on Train_FuelConsumption.csv,
- save your trained model in joblib format,
- and test your model on Test_Simul_FuelConsumption.csv.

To save your model, use:

```
from joblib import dump, load
dump(pipe, "trained_modelXX.joblib")
```

2- **Test_solution.py**.
This file must contain **only the code required to load the trained model and test it**. Use the template below and complete only the X and y extraction:

```
"TESTING SOLUTION"
from joblib import load
import pandas as pd

modeljob_test = load("trained_model_XX.joblib")

test = pd.read_csv("test_FuelConsumption.csv")
X_test = ……    #fill it
y_test = ……    #fill it

y_pred = modeljob_test.predict(X_test)
print("Final score =", r2_score(y_test, y_pred))
```

This ensures that I can later replace Test_Simul_FuelConsumption.csv with my private Test_FuelConsumption.csv.

3- The **trained_model_XX.joblib**
This is your saved model. Replace XX with your FirstNameLastName**.**

Evaluation of the proposed solution

- Your model will be evaluated using the r2_score tested on Test_ FuelConsumption.csv.
- **Model performance (80%)**
  Your model will be evaluated using the r2_score on Test_FuelConsumption.csv. (based on your *Test_solution.py*)
  - The performance score will be obtained using a normalized score

$$S_i = \frac{R2_i - R2_{\text{baseline}}}{R2_{\text{best}} - R2_{\text{baseline}}}$$

Where
- R2_baseline is fixed to 0.9 and
- R2_best is the best student's R^2 on hidden test

The final grade for the model performance is:

$$\text{Performance Grade (\%)} = 50 + 50 \times S_i$$

Additional rules:

- R2 scores will be rounded to three decimal places.
- R2 should be ranged between 0 and 1. R2 score above 1is not accepted
- If your *Test_solution.py* script does not run, the model performance score is zero.

- **Code quality and originality (20%)**

The Solution.py file will be used:
- to check minimal academic integrity,
- to understand your methodology.

Submission format

Submit the three files in a single ZIP archive named:
FirstNameLastName_Competition.zip