

# MACHINE LEARNING - MILESTONE 2 - SUPERVISED LEARNING

---

Francisco P. Romero - University of Castilla La Mancha

November 15, 2022

Based on aspects of building location and construction, your goal is to predict the level of damage to buildings caused by the 2015 Gorkha earthquake in Nepal. The data was collected through surveys by Kathmandu Living Labs and the Central Bureau of Statistics, which works under the National Planning Commission Secretariat of Nepal. We're trying to predict the ordinal variable *damage\_grade*, which represents a level of damage to the building that was hit by the earthquake. More Info about the competition:

<https://www.drivendata.org/competitions/57/nepal-earthquake/>

The dataset mainly consists of information on the buildings' structure and their legal ownership. Each row in the dataset represents a specific building in the region that was hit by Gorkha earthquake. There are 39 columns in this dataset, where the *building\_id* column is a unique and random identifier. The remaining 38 features are described in <https://www.drivendata.org/competitions/57/nepal-earthquake/page/136/>. Categorical variables have been obfuscated random lowercase ascii characters.

You can use as base of your work this Google Colab notebook

<https://colab.research.google.com/drive/16haW1I5j2z9t--FBMDG91uIp0S7bv-Ex?usp=sharing>

## Deliverables

- a report (strong limit 10 pages following the Springer Nature Template<sup>1</sup>) describing the following: the used techniques and their parametrization and the results obtained including the DrivenData user.
- the google colab notebook (.ipynb file) with the code to obtain the best results on the competition.
- the submission file of the best results.

**Deadline: Dec 15th**

---

<sup>1</sup><https://www.overleaf.com/latex/templates/springer-nature-latex-template/gsvvftmrppwq>

## Evaluation

- (2) Report/Code Presentation
- (2.5) Baseline: test with basic regression techniques (kNN, Decision Trees, ) with simple optimization.
- (2.5) Optimization: Use of GridSearch/ to optimize some parametrization of complex techniques (Ensembles, etc.)
- (1.5) Extensions: Introduce variations and lines of improvements
- (1.5) Competition Result.