

Thesis subject: Towards a better evaluation of machine learning models

Laboratory: LIMOS, Clermont-Ferrand, France

Context:

There are many machine learning algorithms whose advantages and disadvantages are generally well known. However, it remains difficult to determine which algorithm to use for a specific dataset, since algorithm user guides typically include only the simplest dataset characteristics (number of attributes, number of classes, etc.). In reality, dataset characteristics are far more complex and have a significant impact on algorithm performance. This PhD project aims to better identify the characteristics for which a machine learning algorithm (unsupervised at first, supervised at a later stage) would achieve good or poor performance.

Topic:

The objectives of the PhD are as follows:

- Propose formulas and define metrics to characterize datasets. Measuring these characteristics has many advantages. For example, they can help researchers better present the strengths and weaknesses of new classification algorithms.
- Associate a type of dataset, defined by its characteristics, with a set of classification algorithms. The ultimate goal is to create user guidelines that, in the context of an applied problem with a specific dataset, help to select the most appropriate algorithms.
- Propose methods to generate new datasets according to the properties a user wishes to investigate. Since existing datasets may not be sufficient to analyze certain characteristics, it is also necessary to develop algorithms that generate synthetic data to fill gaps where datasets with specific characteristics are lacking.
- Study the robustness of models when a percentage of errors is added to the data (in the form of outliers or noisy labels).

Funding: This PhD topic is eligible for doctoral funding from the French Ministry of Higher Education and Research. Candidates must rank within the top third of their cohort to qualify for this type of funding.

Required Education and Skills: Graduate (Master's degree, BAC+5) in computer science, with knowledge in data science and an interest in mathematics.

Application: The application must include a CV, as well as transcripts from the fourth year of higher education (BAC+4) and those already obtained in the fifth year (BAC+5). The application deadline is **April 6, 2026**.

Contacts: Violaine Antoine violaine.antoine@limos.fr, Sébastien Salva sebastien.salva@limos.fr