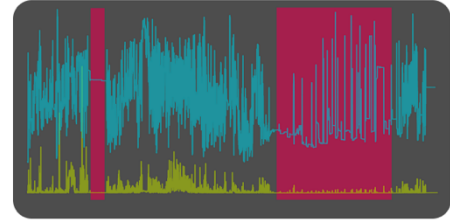# Data & Web Mining Lab Project

Claudio Lucchese   claudio.lucchese@unive.it

Lab Project

Child Mind Institute — Problematic Internet Use
Relating Physical Activity to Problematic Internet Use

- **Prediction Task:**
  https://www.kaggle.com/competitions/child-mind-institute-problematic-internet-use/

- **To be delivered:**
  - One or more notebooks (and possibly other python files) and a 5-page pdf report discussing your critical choices. No more than 5 pages! The pdf report can be the export of a notebook.
  - Project is discussed at the oral exam
  - Deadline: No deadline, just do it within 1 year from now. Upload to moodle and write me an email.

Lab Project

- **Work to be done:**
  - Prepare your data
  - Compare at least two prediction methods, after fine-tuning their parameters
    - One method from scikit learn
    - One method beyond scikit-learn, e.g., with LightGBM or XGBoost or CatBoost
  - Evaluate through Kaggle if possible
  - Investigate importance of features
  - Investigate instances with the most correct predictions and with the most wrong predictions. Do they have some peculiarities? Any strange feature distribution?
  - Can be done in groups, multiply the work by the size of the group

- **Evaluation:**
  - Quality of the report
  - Quality of the code (I'll ask about every single line of code)
  - Depth of the analysis
  - Project can be insufficient (the exam is not passed),
    or it contributes to 30% of the final mark