Master Internship Position

1 Information about the internship

- **Supervisors:**

    - Nour El Mawas (https://nour-elmawas.com/)
    - Germain Forestier (https://germain-forestier.info)

- **Location:** UHA/IRIMAS EA 7499, Mulhouse, France

- **Duration:** 6 months (starting from February or March 2023)

- **Keywords:** Machine Learning, content personalization, Technology Enhanced Learning

2 Context

The ANR COPCOT project involves members from Pixel team (CREM), MSD Team (IRIMAS), Trigone team (CIREL), and France-IOI association. Our aim is to understand learners' skills-building in the Technology Enhanced Learning platform Quick-Pi, and to facilitate skills-building through content and feedback personalization.

3 Goals

The goal of this internship is to analyze learners' data from the Quick-Pi platform (https://quick-pi.org/contenu.html) in order to identify relevant learner behaviors and extract data indicators contributing to the skills construction in programming among learners.

In particular, the candidate will work on the following tasks:

- **Learners' data analysis.** Analysis of trace logs data from users' interactions help us to better understand their learning process, distinguish groups of learners, and predict learners' success in each Quick-Pi exercise. Trace logs data can be seen as time series, hence this task will start by reviewing existing Machine Learning algorithms [1] used for time series analysis. Then the candidate will identify and adapt these algorithms to our sequential data in order to predict the learners' success in each exercise

on the Quick-Pi platform. The implicit data provided by Quick-Pi logs (like time spent on an exercise, navigation logs, code lengths, etc.) will be used. Once the pertinent algorithm to predict students' success is found, indicators can be found. Existing learners' data analysis techniques such as knowledge tracing [2, 3] will be investigated to predict the success of students in each exercise.

- **Indicators identification.** The candidate will analyze which logs in the selected algorithm of the previous task were used by the algorithm itself to predict the success. This will help us to identify indicators. We believe that if the selected algorithm successfully predicts learners' success, we should be aware of which specific log lines contributed to this prediction the most. For example, we can find that the time spent on an exercise may have a correlation with the learners' success in a specific exercise. These indicators are crucial to understand skills-building and learners' difficulties. Features selection techniques [4] and explainability techniques [5] will be investigated in order to identify pertinent indicators for skills-building.

4 Profile of applicant

The candidate must fit the following requirements:

- Registered in Master 2 or last year of Engineering School (or equivalent) in **Computer Science**

- Good skills in **Python programming** are mandatory

- Good skills in **Machine Learning** are required

- Interest, knowledge and/or a first experience in **Technology Enhanced Learning (learning programming in particular)** will be appreciated

5 Research environment

The proposed internship will be part of the ANR COPCOT project starting in 2023 for 5 years.
Hence, there is a great opportunity to continue with a PhD on the same topic/project.

6 Application

For further information or for applying, candidates should send the following documents to nour.el-mawas@univ-lorraine.fr:

- A CV

- Academic records

- Personal projects (e.g. github repo)

- A motivation letter

References

[1] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: a review," *Data mining and knowledge discovery*, vol. 33, no. 4, pp. 917–963, 2019.

[2] M. Sao Pedro, R. Baker, and J. Gobert, "Incorporating scaffolding and tutor context into bayesian knowledge tracing to predict inquiry skill acquisition," in *Educational Data Mining 2013*, Citeseer, 2013.

[3] S. I. Ramírez Luelmo, N. El Mawas, and J. Heutte, "Existing machine learning techniques for knowledge tracing: A review using the prisma guidelines," in *International Conference on Computer Supported Education*, pp. 73–94, Springer, 2022.

[4] V. Vijayalakshmi and A. Prakash, "Developing an optimized feature selection process for designing efficient content management system using educational data," *International Journal of Advanced Research in Science Technology (IJARST)*, vol. 7, no. 1, pp. 15–24, 2020.

[5] R. Alamri and B. Alharbi, "Explainable student performance prediction models: a systematic review," *IEEE Access*, vol. 9, pp. 33132–33143, 2021.