Summary report

August 2021

This paper aims at summarizing the deliverables associated to the outcomes of the research internship on *Predicting students' performance in online courses using multiple data sources* performed by Mélina Verger (Paris-Saclay University) between March and August 2021.

The objective of the internship was to address students' performance prediction with the aim of answering the following research questions:
1. To what extent using multiple data sources on learners increases prediction performance?
2. Which learners' characteristics are most related to higher academic performance?

Moreover, one of the contributions of this project was to explore a generic approach to create a course and collect data from it. This project may be thus useful for anyone interested in creating a challenge in Educational Data Mining (EDM) and in extending our preliminary work to reach conclusive findings with larger datasets and additional experiments.

1 Report

The report of the internship can be found at `https://github.com/melinaverger/ed_project/blob/main/documents/report.pdf`.

2 Documentation

The following link `https://github.com/melinaverger/ed_project/blob/main/documents/documentation.pdf` will direct you to the documentation which describes how the course was implemented and provides some guidelines to create a course and to collect data on The Open University e-learning platform.

3 GitHub repository (`ed_project`)

The repository `https://github.com/melinaverger/ed_project` contains the codes to reproduce the experiments or to exploit for one's own usage.

4 Codalab challenge

We created a toy challenge with the data we collected on Codalab platform at `https://competitions.codalab.org/competitions/34484?secret_key=6f6855ad-80a7-49ca-aa79-b7e8926c8e00`. The associated GitHub repository, `ed_challenge`, is available at `https://github.com/melinaverger/ed_challenge`. Most of the results are based on a small number of instances, stressing the limitations of the study.