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 (INAOE / Paris-Saclay University) between March and August 2021.

1 Introduction

1.1 Objective

The objective of the project was to address the students' performance prediction task 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?

1.2 Contribution

In addition to the results obtained (see subsection 2.1), one of the contributions of this project was to explore a generic approach to create a course and collect data from it. This may be useful for anyone interested in working with educational data and performing relative experiments in Educational Data Mining (EDM).

While our initial goal was to design a real challenge on the collected data, we were not able to collect enough samples and therefore we are only providing a toy challenge (see subsection 2.4). Nevertheless, we provide as deliverable the guidelines, code and an implementations so that these deliverables comprise useful resources for organizing the related challenge as future work.

2 Deliverables

2.1 Report

The report of the project can be found at `https://github.com/melinaverger/ed_project/blob/main/documents/report.pdf`. The motivation and the background of the project, the methods used, the experiments performed and the results obtained are thoroughly explained follwing the ICML 2021 latex template[1]. At this time, most of the results are based on a small number of instances.

2.2 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[2].

---

[1] `https://www.overleaf.com/latex/templates/icml2021-template/dsftnbmjgyhv`
[2] `https://www.open.ac.uk`

2.3   GitHub repository

The `ed_project` repository `https://github.com/melinaverger/ed_project` contains the codes to reproduce the experiments or to exploit them for one's own usage. It is the main location for all the project resources. Usage information are presented in the repository.

2.4   Codalab challenge

We created a toy challenge with the data we collected on the 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` and contains the materials to implement the challenge.