



Dropout Prediction in MOOCs

PID: 2021\DSAI\010

The project is on analyzing the data generated on the NPTEL platform, one of the biggest MOOCs worldwide, and predicting the dropouts, participants who leave the course before completing it. More than a million participants register for hundreds of courses on NPTEL every semester. One of the key challenges while running an educational program at such a massive scale is to minimize the number of dropouts. Some recent studies develop machine learning-based methods to predict dropouts. However, these studies are limited as their validation is done on small/public datasets. This project's primary goal is to develop machine learning-based algorithms to predict dropouts and interpret the predictions that can lead to possible interventions to minimize dropouts and enhance the learning experience of enrolled students.

Task to be assigned to the intern: Data preprocessing, reading papers on existing methods, writing code for machine learning models, and articulating results.

Learning outcomes: At the end of internship, the student would develop the following skills

- Hands-on experience of working with real-world data.
- Challenges associated with data preprocessing to interpret the results will season the student to work on the complete pipeline of developing ML-based methods.
- Developing machine learning models for extremely large datasets. Student will develop working knowledge of the ML-models at such large scale.

If successful, the project might lead to a quality publication in top-tier ML/Data Mining conference/journal. Any improvement in the state-of-the-art methods carries a huge significance to the e-learning community. The contributions will get exceptional visibility and help the student secure admission into graduate-level research programs in top universities globally.

Duration: 3-6 months

Programming: At least one programming language (preferably python)

Other skills: Familiarity with machine learning. Knowledge of deep learning will be a plus

Prerequisite courses: Any one of machine learning, data mining, pattern recognition or equivalent

Number of interns required: 1 - 2

References:

1. Wei Wang, Han Yu, and Chunyan Miao. 2017. Deep Model for Dropout Prediction in MOOCs. In Proceedings of the 2nd International Conference on Crowd Science and Engineering (ICCSE'17). Association for Computing Machinery, New York, NY, USA, 26–32. DOI:<https://doi.org/10.1145/3126973.3126990>
2. Moreno-Marcos, P.M., Munoz-Merino, P.J., Maldonado-Mahauad, J., Perez-Sanagustin, M., Alario-Hoyos, C. and Kloos, C.D., 2020. Temporal analysis for dropout prediction using self-regulated learning strategies in self-paced MOOCs. Computers & Education, 145, p.103728