

Ouroboros: early identification of at-risk students without models based on legacy data

Title, Authors, Source, Year

Title: Ouroboros: early identification of at-risk students without models based on legacy data

Author: Hlosta, Martin; Zdrahal, Zdenek and Zendulka, Jaroslav

Source: <https://core.ac.uk/display/83955564?source=2>

Year: 2017

Level 1 Summary:

Ouroboros project focuses on identifying students at the risk of failing their course, with the absence of data from previous running's of the course which are usually the de facto method of training models for such systems. The system uses information from already submitted assessments instead in the running presentation of the course. Four types of data are used students' demographic information (age, gender, etc.), students' interactions with the VLE system, information about students' date of registration and a flag indicating student assessment submission. The expectation that the behaviour of learners who about to submit will follow a similar pattern to those who have already submitted and differ from students who will not submit. The study chose models that support probabilistic predictions; this included Logistic Regression (LR), Support Vector Machines (SVM), Random Forest (RF), Naive Bayes (NB), and the Tree Boosting XGBoost. A problem that had to be overcome was imbalanced datasets. The use of cost sensitive learning and other approaches help to mitigate this issue. The models predict the likelihood a student will submit coursework each day before the deadline. The closer the deadline, the higher precision each model receives.

Why did I read this paper?

I read this paper to understand an alternative solution on creating an early warning system with limited previous course data and student data. This approach can help me understand the alternative solutions that could be used in my own research.

Personal view:

This approach has been an interesting way to get around the issue of lack of course data and student data. The paper was useful to understand the challenges of a similar approach and a base accuracy a similar system might have.

What problem does this paper address?

The paper solves the issue of lack of previous course data and student data to train models for early warning systems for at risk students.

Is it an important problem?

This is a very important problem currently. Institutions are keen to give more support to students that are possibly at risk to prevent students dropping out. A lot of new courses don't have the previous data to develop ml models and 1st year students don't have the past academic data to help train models. This system can help solve some of these problems to make sure all courses can provide early warning systems for at risk students.

What is the significance of the result and its solution?

The results show that models can be trained using current in course data and still provide precise results which can be utilised in similar systems.

What are the claimed novel contributions of the paper?

The utilisation of information about already submitted assessments to make predictions instead of legacy data.

What previous work is the basis for this research?

The study is based on previous research by Open university which identified students who were at risk using previous running's of the course. A. Wolff, Z. Zdrahal, D. Herrmannova, J. Kuzilek, and M. Hlosta. Developing predictive models for early detection of at-risk students on distance learning modules.

The dataset OULAD has also been used during this study. J. Kuzilek, M. Hlosta, and Z. Zdrahal. Open university learning analytics dataset.