OU Analyse: analysing at-risk students at The Open University

Title, Authors, Source, Year

Title: OU Analyse: analysing at-risk students at The Open
University

Author: Kuzilek, Jakub; Hlosta, Martin; Herrmannova, Drahomira; Zdrahal, Zdenek;
Vaclavek, Jonas and Wolff, Annika

Source:
http://oro.open.ac.uk/42529/1/__userdata_documents4_ctb44_Desktop_analysing-at-risk-students-at-open-university.pdf

Year: 2015

Level 1 Summary:

OU Analyse project is used to predict students at possible risk as early as possible
within a course. The study uses two type of data about the students; demographic data
(Sex, New/Cont Student, Education previously) and interaction of the VLE (Virtual
Learning Environment). Using the data, 4 predictive models are created (Bayesian
classifier, Classification and regression tree (CART), k Nearest Neighbours (k-NN)
with demographic/static data, k-NN with VLE data). Each of the 4 models predict the
students likelihood to be "at risk". If 2> models predict the student is at risk, the
student will be noted as a at risk student, otherwise student is not seen at risk. Weekly
predictions are sent to the course coordinator who can contact the relevant student for
additional support. The predictions have a precision of about 50% at the beginning of
a semester to over 90% by the end. The recall rate decrease from 50% to 30% in the
same period.

Why did I read this paper?

I read this paper to understand how a large (mainly online) university is able to
develop a early warning system and the type of methods which are used. By seeing
which models work well, this will hopefully help inform myself and find out which
models I should look at using in my project.

Personal view:

In my opinion I thought this was overall an interesting paper. It help me understand
why which different models have been used in the study and how they have been
implemented. I feel using multiple different models/strategies might be key to solving
my own research question.

What problem does this paper address?

The paper solves the issue of students not getting the support early enough when they
are at risk of falling behind and dropping out. With class sizes increasing it can be
hard to keep up which students are doing well and others that are not.

Is it an important problem?

This is a fairly important problem right now for all educational institutions. With class sizes increasing and teaching being held online, there is more of a requirement to developing a system to allow course coordinator to give support if required to "at risk" students.

What is the significance of the result and its solution?

By using multiple different models, the strength of each model can be utilized to get the most precise predictions.

What are the claimed novel contributions of the paper?

The study has been worked rolled out to 10+ courses and is currently showing promising results.

What previous work is the basis for this research?

Using machine learning to identify students at risk has been used in a number of publications including (Arnold & Pistilli, 2012; Baradwai & Pal, 2011; Huang & Fang, 2013; Kabra & Bichkar, 2011; Pandey & Sharma, 2013; Romero, et al., 2013; Wladis, et al., 2014; Wolff, et al., 2014; Hlosta, et al., 2014).

This paper is based on the previous research of (Wolff & Zdrahal, 2012; Wolff, et al., 2014; Wolff, et al., 2013; Wolff, et al., 2013; Hlosta, et al., 2014). This approach utilised decision trees using VLE data and scores in assessment to predict students behaviour.