Old Dominion University

CS 773

Data Mining and Security

Course Project

Authors:
Grant Atkins
David Haslam

July 31, 2017

Contents

1 Executive Summary

The goal of this course project, originally stemmed from Open University Learning Analytics dataset, is to identify ‘at-risk‘ students provided [2].

More needs to be done in this section...

2 Introduction

With analytics becoming an integral part of Learning Management Systems (LMS), Universities can analyze their data to intervene and aid ‘at-risk‘ students. This can be done by detecting failing or struggling students earlier on in their courses. These detections can later be used to predict ‘at-risk‘ students for future intervention. Analysis of demographic information for students may also prove useful to detect environments or character attributes that identify a struggling group.

The Open University (OU) in the United Kingdom, offered distance learning courses with its virtual learning environment (VLE) from 2013 to 2014. Students interacted with the VLE and OU then collects information such as: page names, clicks, and times connected to the website. Students are also required to register for these courses earlier on, which provides: registration date, unregistration date, class types and semester. Many students may end fail or withdraw from the VLE courses. The datasets that OU provided can be used to provide insight on 2013 to 2014 semester students.

3 Problem Statement

When analyzing students to observe for ‘at-risk‘ students it may seem easily done by looking at simply grade scores, however, there could be independent factors present in a student’s personal history or personal accomplishments that cause a student to be ‘at-risk.‘ The goal of this paper is discuss and find data mining techniques that aid in the detection and prediction of ‘at-risk‘ students.

4 Solution Methodology

Finding a solution for this project was based on the data being observed. We first attempted to look at each dataset, each csv provided, individually and then find attributes that would be the best indicators to determine the 'final_result' of students. Later we would merge the datasets togethers and then perform the following steps depending on the dataset:

1. Find the best attributes for the data provided, often through information gain for feature selection or PCA for clustering
2. Test machine learning techniques such as:
   (a) Decision trees
   (b) Naive bayes
   (c) Bagging
   (d) KNN
   (e) Kmeans
3. Cross validation for measuring effectiveness

This paper first talks about the individual datasets and then the results of the merged datasets

4.1 Individual data

4.2 Merged data

5 Experimental setup and data used

There are many different ways to experiment on OU learning analytics dataset. There are many interactions between the different sets of data, such as demographic information can be linked to course assessments for each student as shown in Figure 1.

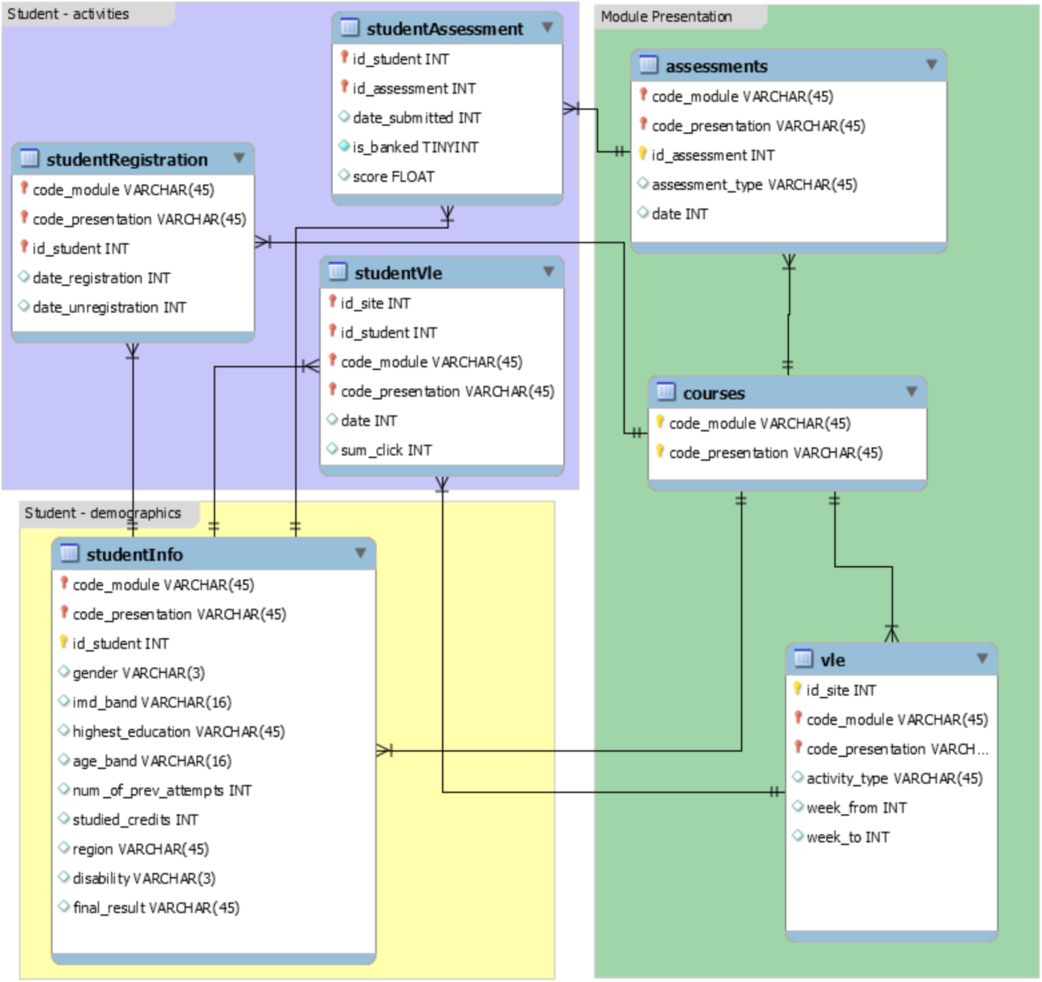

Figure 1: OU learning analytics dataset relations

6 Results

7 Conclusions

— End

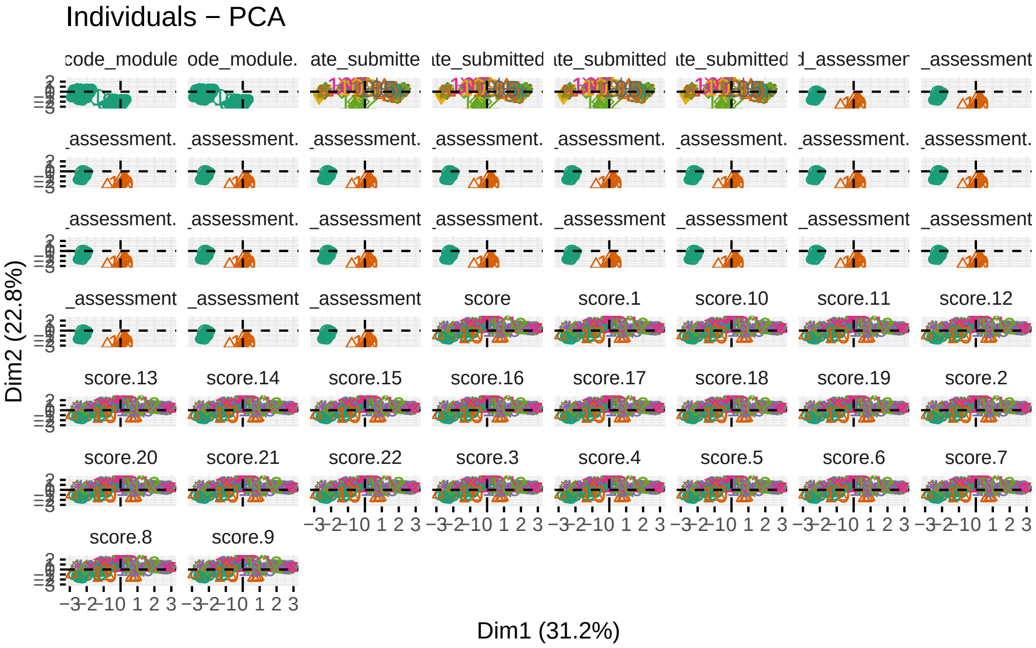

Figure 2: PCA of student assessment attributes

References

[1] Help on BibTeX entry types. `http://nwalsh.com/tex/texhelp/bibtx-7.html`. Accessed: 2015-03-12.

[2] Hlosta M. Herrmannova D. Zdrahal Z. Kuzilek, J. and A. Wolff. test. `http://www.laceproject.eu/publications/analysing-at-risk-students-at-open-university.pdf`, March 2015.