Open University Intervention to increase course completion rates (by preventing students from Withdrawing)

Introduction

This document outlines the proposal for the Capstone project for successful completion of the Data Science Intensive program. This proposal outlines the problem, its intended client and how it is relevant. Some information about the source and acquisition of data is provided. Finally, an attempt is made to discuss how the problem will be approached and expected solution. The final deliverables at the end of the course (capstone submission) will conclude the proposal

Problem statement

The initial idea for the project comes from thinking about motivation levels for online courses. In more detail, it leads to thinking about who are the students, gender and age distribution, background. How many of these students withdraw from the courses and is it possible to stop the trend?

To sum up, the problem is to classify students who withdraw from self-learning/online courses so that earlier intervention can prevent them from withdrawing from the course

Relevance to the Client

The client will be Open University who can use the classification to increate the proportion of students who withdraw from their courses. This can be applied to other courses too with a high degree of self-learning such as online courses. It will not only make the courses more relevant and useful to the students but also increase their success rate.
There are assumptions and theories related to the profile of these students and influencing factors in their decision to withdraw. hTis solution will justify these factors or provide a new insight based not on intuition but on concrete data analysis.

Data

The data for the project has been made available by Open University, Uk and is available for download at:

https://analyse.kmi.open.ac.uk/open_dataset

The website also provides a database schema along with details related to each table.

While initially it is not sure which tables will be used for analysis, it is a good starting point.

For the purposes of the project I hope to use the following:
studentInfo.csv
studentVle.csv
studentAssessment.csv

Approach to problem

The three steps that follow the problem to achieve a suitable solutions will be:
1. Data cleanup and initial analysis: Make sure data is not incomplete and form an initial map of data
2. Exploration and analysis: Explore the data with a focus on student profile as well as any leads about students that withdraw. Analyze data to create a profile and weightage about factors that can give a preview to students who are “at risk”
3. Machine learning: Use classification algorithm that can be used to create the profile and hence allow educators to intervene and make sure more students complete the course

At this point, the algorithm that will be used is not determined.

Deliverables

Initial submission: Project proposal
After initial analysis: Data story, Milestone report
Final submission:
1. Code for the project
2. Slide deck for final presentation
3. Final project report explaining the problem, analysis of data, approach to solution, and findings. Some idea for further research as well as limitations. Conclude with recommendations for your client on how to use your findings.

3 A slide deck or a blog post which presents your analysis to your clients (e.g. non-technical and business teams) in an easy to understand, but compelling way.