Kim Pham
Springboard
Capstone Project
May 2017

Capstone Milestone Report

What is the problem?

The use of technology is not limited to our personal and work life. It is also being used in the K-12 education space. Education technology is a relatively new field worth exploring because incorporating technology into the classroom, such as apps dedicated to personalized learning and gamification of learning, is changing the traditional teacher-student model.

The Project

I am curious about the extent which technology is helping students achieve success. I specifically want to explore the effectiveness of education technology in helping students achieve high scores. In completing this project, I hope to understand how technology does or does not help in the education space by utilizing and learning pandas, statistics, machine learning, and data story telling.

Who is the client?

This capstone project may be of interest to anyone interested in using, investing in, or building education technology software aimed to improve student learning.

A Deep Dive into the Data Set

For this project, I will be exploring the data set found here: https://www.kaggle.com/aljarah/xAPI-Edu-Data (Date: 2016-11-8).

"This is an educational data set which is collected from learning management system (LMS) called Kalboard 360. Kalboard 360 is a multi-agent LMS, which has been designed to facilitate learning through the use of leading-edge technology. Such system provides users with a synchronous access to educational resources from any device with Internet connection."

"The dataset consists of 480 student records and 16 features."

```
['gender',
 'NationalITy',
 'PlaceofBirth',
 'StageID',
 'GradeID',
 'SectionID',
 'Topic',
 'Semester',
 'Relation',
 'raisedhands',
 'VisITedResources',
 'AnnouncementsView',
 'Discussion',
 'ParentAnsweringSurvey',
 'ParentschoolSatisfaction',
 'StudentAbsenceDays',
 ‘Class']
```

Attribute Details

1 Gender - student's gender (nominal: 'Male' or 'Female’)

2 Nationality- student's nationality (nominal:’ Kuwait’,’ Lebanon’,’ Egypt’,’ SaudiArabia’,’ USA’,’ Jordan’,’ Venezuela’,’ Iran’,’ Tunis’,’ Morocco’,’ Syria’,’ Palestine’,’ Iraq’,’ Lybia’)

3 Place of birth- student's Place of birth (nominal:’ Kuwait’,’ Lebanon’,’ Egypt’,’ SaudiArabia’,’ USA’,’ Jordan’,’ Venezuela’,’ Iran’,’ Tunis’,’ Morocco’,’ Syria’,’ Palestine’,’ Iraq’,’ Lybia’)

4 Educational Stages- educational level student belongs (nominal: ‘lowerlevel’,’MiddleSchool’,’HighSchool’)

5 Grade Levels- grade student belongs (nominal: ‘G-01’, ‘G-02’, ‘G-03’, ‘G-04’, ‘G-05’, ‘G-06’, ‘G-07’, ‘G-08’, ‘G-09’, ‘G-10’, ‘G-11’, ‘G-12 ‘)

6 Section ID- classroom student belongs (nominal:’A’,’B’,’C’)

7 Topic- course topic (nominal:’ English’,’ Spanish’, ‘French’,’ Arabic’,’ IT’,’ Math’,’ Chemistry’, ‘Biology’, ‘Science’,’ History’,’ Quran’,’ Geology’)

8 Semester- school year semester (nominal:’ First’,’ Second’)

9 Parent responsible for student (nominal:’mom’,’father’)

10 Raised hand- how many times the student raises his/her hand on classroom (numeric:0-100)

11- Visited resources- how many times the student visits a course content(numeric: 0-100)

12 Viewing announcements-how many times the student checks the new announcements(numeric:0-100)
13 Discussion groups- how many times the student participate on discussion groups (numeric:0-100)
14 Parent Answering Survey- parent answered the surveys which are provided from school or not (nominal:'Yes','No')
15 Parent School Satisfaction- the Degree of parent satisfaction from school(nominal:'Yes','No')
16 Student Absence Days-the number of absence days for each student (nominal: above-7, under-7)

The students are classified into three numerical intervals based on their total grade/ mark:
Low-Level: interval includes values from 0 to 69,
Middle-Level: interval includes values from 70 to 89,
High-Level: interval includes values from 90-100.

What kind of cleaning and wrangling did you need to do?
- The Data set is pretty clean. There are no Nulls.

Limitations

What are its limitations i.e. what are some questions that you cannot answer with this data set?
- It's important to keep in mind the attributes available and not available in this dataset. For example, we have information on 'Raised hand' which might illustrate classroom participation outside of using technology, but that might not be enough to describe the in-person classroom dynamic, culture, or peer and teacher relationships. That might require more data acquisition to fully understand other elements which might affect student success in the classroom.

Are there other datasets you can find, use and combine with, to answer the questions that matter?
- I want to stick with this one dataset. I think there's a lot of content here already, and as my first project, I want it to be as simple as possible to start with.
- With that said, I will continue to build off what Dan did with this same dataset here: https://www.kaggle.com/dan195/classification-of-student-marks
- Dan has some great data visualization and exploration, classification, and calculations using Linear and Non-Linear SVCs.

Final Approach

Based on these findings, what approach are you going to take? How has your approach changed from what you initially proposed, if applicable?

- I will be applying different models, such as linear and logistic regressions, to see if we can improve the accuracy of Dan's model predictions.