Geo-Based Prediction of Academic Performance

Topic outline

Star Chen (sc7828) - *Computer Science*
Scott Ye (my2152) - *Data Science (CS Concentration)*

Supervision

- Ratan Dey (rkd230@nyu.edu)
- Promethee Spathis (ps147@nyu.edu)

Context

Due to COVID-19, instructors worldwide have shifted the teaching mode to online in response to the stay-at-home orders. Learning management systems (LMS) such as Canvas, Moodle, and Blackboard have been adopted by instructors, which enables students to complete coursework regardless of location or time availability. However, feedback from educators shows concerns about this new learning format. Notably, because of the lack of face-to-face interactions, it becomes difficult for instructors to assess student's learning progress. Therefore, a reliable technique for academic performance prediction is needed within the online learning context so that instructors can carry out pre-intervention measures to prevent academic failures.

Considerable state-of-the-art machine learning algorithms have been successful in the prediction of students' online academic performance [1]. However, current research mainly focuses on students' study behaviors and demographics for the prediction of students' academic performance, while factors including students environment have been largely neglected [2]. Meanwhile, most of the studies that use students environment as a factor, such as [3], concentrate on the virtual environment, in which case research on physical environment is lacking.

The goal of this project is to investigate the significance of students' geolocation, a representative feature of physical environment, and to predict students' academic performance based on students' geolocation and study behaviors. We will analyze data provided by one popular LMS, Moodle, to help our prediction. The prediction will allow instructors to derive a list of possible pre-intervention measures.

Objectives

The overall objective of this project is to answer the following research questions:

1. What is the impact of stay-at-home orders on students' academic performance?
2. Can academic performance be predicted based on students' study behaviors?

Upon completion of this project, we expect that we will have the following accomplishments:

1. Design of a learning algorithm to predict students' academic performance based on their geolocation.
2. Identification of correlations between students' academic performance and their study behaviors.

To realize these objectives, we will analyze the log reports, collected from a widely used LMS called Moodle, of 11 online quizzes given during Fall 2020 and 2021 to a cohort of 250 undergraduate students. The log reports include the following information for each student who took the quizzes:

– The dates the quiz was started and completed.

– The time taken to complete the attempt.

– The grade and the details of the points per question for the finished attempts.

– The type of device the student was logged to.

– The IP address of the device the student was logged to.

We will derive the students' geolocation from their IP addresses, and we will identify various study behaviors and find correlations between student location data, time, and date. In particular, our research will contain the following stages:

1. Conduct a literature review for a better understanding of the existing approaches.

2. Experiment with different approaches (Random Forest, AdaBoost, Decision Tree, Multi-Layer perceptron...) and determine the ideal method for prediction.

3. Train the model that predicts academic performance based on geolocation and other factors; conclude on the correlation between academic performance and their study behaviors.

4. Evaluate the learning algorithm with reports from Moodle or data sets from other sources.

5. (If time permits) explore and implement several advanced machine learning models and optimize the algorithm.

References

[1] T. Panagiotakopoulos, S. Kotsiantis, G. Kostopoulos, O. Iatrellis, and A. Kameas, "Early dropout prediction in moocs through supervised learning and hyperparameter optimization," *Electronics*, vol. 10, no. 14, p. 1701, 2021.

[2] A. Abu Saa, M. Al-Emran, and K. Shaalan, "Factors affecting students' performance in higher education: A systematic review of predictive data mining techniques," *Technology, Knowledge and Learning*, vol. 24, no. 4, p. 567–598, 2019.

[3] J. A. Lara, D. Lizcano, M. A. Martínez, J. Pazos, and T. Riera, "A system for knowledge discovery in e-learning environments within the european higher education area – application to student data from open university of madrid, udima," *Computers amp; Education*, vol. 72, p. 23–36, 2014.