Predicting Undergraduate Students' Final Exam Scores in Online Synchronous History Class

Rebecca, Jingxuan, Mark

Literature Review

- Educational data mining can be used to identify students' behavioral patterns and predict their grades (Romero & Ventura, 2010).
- Previous studies on student performance prediction used analytic methods such as Bayesian Network, Association rule, PCA, Classification, etc., to build prediction model (Zhang et al., 2017)
- Identify critical factors to improve prediction accuracy and the accepted range of prediction accuracy ranged from 75% to 95% (Asif et al., 2014; Villagrá-Arnedo et al., 2016).
- Using machine learning to build a model to predict likelihood that a student would answer a math problem correctly (Beck & Woolf, 2000)

The Problem Statement

What behavioral factors may affect undergraduate students' test scores in the American history course in one semester in the synchronous online settings like Zoom & Canvas?

Population: Undergraduate students in the American history class

Duration: One semester

Location: Zoom and Canvas

Objectives

- Define behavioral factors that affect final test scores
- Collect, clean, and analyze data
- Build a model to predict final test scores
- Build clusterings of students to group them into different review sessions

Data Collection

- **Participation**
  - Number of hand-raising in Zoom live session
  - Number of discussion posts in Canvas
  - Number of absence days
- **Learning**
  - Number of visits to class resource page in Canvas
  - Completion of resource watch
  - Number of studying hours
  - The number of office hour attendance

Data Collection

- **Performance**
  - Online assignments completion
  - Average essay scores
  - Average previous test scores
- **Course requirement**
  - Elective or Mandatory course

Data Cleaning

- Substitution of null value (median/mean/mode)
- Detection of outlier
- Standardization of variables
- Transformation of data type (categorical → numerical)

Data Analysis Plan

- **Description Analysis and Plots**

  - Illustrate the relationship between the average test scores with each variables

- **Explore the correlation between the scores and each variable**

  - To explore which factors have obvious impact on the test score.

- **Structure Discovery: Clustering Analysis**

  - Grouping students by the collected variables

- **Predictictive method: Classifiers → Logistics Regression Analysis**

  - Use Logistics Regression Analysis to predict students' next history test score. And discuss the relationship between the groups and scores

Expected Results

- The proposed model will achieve high accuracy for predicting student's final exam score.
- Number of absence days is negatively related to the final exam score.
- Critical factors that have greater effect include:
  - Number of hand-raising in Zoom live session
  - The number of office hour attendance
  - Average homework scores (total number of essay scores/number of homework)
  - Average quiz scores (total number of test scores/number of tests)
  - Whether it is a core course or not

Implications

- Offer educators and students more insights on what behavior features could contribute to the performance of the online study
- Suggest future researchers to investigate other behavioral factors that affect undergraduate student's learning in the online learning setting
- Sheds light on an instructional method of assigning students to appropriate learning groups to provide more targeted short-term instructions

Challenges

- Data Accuracy
- Non-exhaustive list of variables
- Data Privacy

Limitations

- Data quantity
- Geographical and cultural exclusiveness

Thank you