Predict Student Performance from Game Play

Yueqi Xu

Dataset: Game Logs From Jo Wilder

- Data were captured from the Jo Wilder online educational game.
- Full Dataset Available on Kaggle.
- The dataset records game log of 23,562 game play sessions, contains information about game setting (full screen or not, music on/off, etc.), gameplay events (click, hover, etc.), and player performance (whether the in-session questions were answered correctly).
- Each game play session in the dataset consists of 18 questions, which are distributed across three checkpoints: 3 questions in the first checkpoint, 10 questions in the second checkpoint, and 5 questions in the third checkpoint. Each checkpoint contains a different set of variables.

Scientific Questions

- What is the optimal combination of features (e.g., time spent on task, number of mouse clicks, etc.) and choice of model for predicting student performance (correctness of answering questions) in a game play session?
- Is there an association between a student's performance on earlier questions and their performance on subsequent questions within the same game play session?
- How can the predictive models for student performance (correctness of answers) during game play sessions be integrated into real-time feedback mechanisms in games to provide timely and personalized feedback to students, thereby enhancing their learning experience and performance in the game?

Method

- The analysis was performed at the level of the checkpoint, and for simplicity, only questions 3, 13, and 18, which are the last question from each checkpoint, were investigated.
- Performance of Logistic regression models and random forest models are compared.
- From EDA: the correctness of questions 3, 13, and 18 are approximately 0.93, 0.27, and 0.95, respectively, which are highly imbalanced. Therefore, models fitted on oversampled data was also considered.
- In general, 4 models are fitted for each of questions 3, 13, and 18:

  - 2 logistic regression models: The correctness of the question vs. the overall correctness of previous questions, adjusting for all other features that appeared in the corresponding checkpoint, with & without oversampling the minority group.
  - 2 random forest models: The correctness of the question given the overall correctness of the previous questions and all other features that appeared in the corresponding checkpoint, with & without oversampling.

Result: Association

- According to the logistic regression models with unbalanced data, for each of question 3, 13, and 18, we have enough evidence from data that there is an association between a student's performance on earlier questions and their performance on subsequent questions within the same game play session.

Question	3	13	18
P-value	1.73e-48	4.63e-71	5.61e-94

Result: Model Performance

- Without oversampling, Both logistic regression (LR) models and random forest (RF) models assigned almost all testing data to the majority group.
- When oversampling was applied to balance the data, both LR models and RF models obtained more reasonable confusion matrices.
- For predictive purposes, I recommend utilizing a random forest model trained on balanced data, with all available features as predictors.

Model	LR ACC	RF ACC
Q3 UB	0.934	0.934
Q3 B	0.712	0.934
Q13 UB	0.732	0.736
Q13 B	0.732	0.785
Q18 UB	0.95	0.951
Q18 B	0.95	0.953

Conclusion 3

- In order to provide timely and personalized feedback to students, one potential application is to leverage this predictive model to continuously monitor the students' actions and performance during the gameplay session.
- A system can be built based on this model to offer feedback and guidance whenever it detects a need for further attention or improvement.
- However, further investigation and development are necessary to establish a more comprehensive and refined feedback mechanism.