Avengers Proposal

## **Topic Selection**

Many topics are wild selected in recent days, for examples, market analysis, fraud detection, health care analysis, recommendation, bio-info mining, etc. However, Higher education teaching method study become more popular due to the rapid development of digital technology highly **impacted** our daily life and **reconstructed** higher education. Even LU leads its own study on **higher education** and also raises a “China and Higher Education Conference 2024” [1]

Moreover, some of our group members had experience in information technology teaching through multi media way. Hence, we decided to figure out the latent relationships between “**student performance**” and “**game play teaching method**”

## **Background**

With the continuous progress of educational technology, the combination of educational content and game mechanism to let students learn in a relaxed and pleasant environment has gradually become an important strategy to improve student participation and learning effect. By **analyzing the behavioral data** of students in games, we can gain insight into students' learning habits, social interactions, and emotional states, providing data support for educators to develop more effective and personalized teaching strategies.

## **Study Aim**

The study aim of this project is to build an effective predictive model to identify the key factors that affect student performance in the game. Specific objectives include:

Data collection and analysis: Collect students' behavior data and related background information in the game, and conduct in-depth analysis.

Feature engineering: Extracting meaningful features from raw data to improve the predictive power of the model.

Model building and optimization: Use data mining techniques to build and optimize predictive models and explore the effects of different algorithms.

Interpretation and application of results: Provide insights into student learning behavior and make data-based recommendations to help educators develop personalized teaching strategies.

## **Data Collection**

We may design a control experiment and collect examination results from students to finish the data collection. But this time consuming way is unacceptable in our group project, thus we try to collect data from website platform Kaggle instead. [2]

## **Data Description**

We use training set, test set, training labels as our original data input. They recording attributes below. [2]

|  |
| --- |
| session\_id - the ID of the session the event took place in  index - the index of the event for the session  elapsed\_time - how much time has passed (in milliseconds) between the start of the session and when the event was recorded  event\_name - the name of the event type  name - the event name (e.g. identifies whether a notebook\_click is is opening or closing the notebook)  level - what level of the game the event occurred in (0 to 22)  page - the page number of the event (only for notebook-related events)  room\_coor\_x - the coordinates of the click in reference to the in-game room (only for click events)  room\_coor\_y - the coordinates of the click in reference to the in-game room (only for click events)  screen\_coor\_x - the coordinates of the click in reference to the player’s screen (only for click events)  screen\_coor\_y - the coordinates of the click in reference to the player’s screen (only for click events)  hover\_duration - how long (in milliseconds) the hover happened for (only for hover events)  text - the text the player sees during this event  fqid - the fully qualified ID of the event  room\_fqid - the fully qualified ID of the room the event took place in  text\_fqid - the fully qualified ID of the  fullscreen - whether the player is in fullscreen mode  hq - whether the game is in high-quality  music - whether the game music is on or off  level\_group - which group of levels - and group of questions - this row belongs to (0-4, 5-12, 13-22) |

## **Methods Applying**

Methods to be applied. [3]

Data acquisition

Data cleaning

Data preprocessing

Feature engineering

Model selection

Model training

Model verification

## **Deep Diving**

Whether the dataset is public for use and has sufficient information to mine.

Why we choose certain data mining method in this study.

How to identify the key attributes influencing the prediction.

What causes the improvement of prediction accuracy.

## **Social Impacts**

Through the implementation of this project, we expect to be able to provide educators with a practical teaching aid model to help them identify the strengths and weaknesses of students in learning through games. The ultimate goal is to advance the development of personalized learning so that every student can get the best learning experience according to their needs. [4]

## **Reference**

1. <https://www.ln.edu.hk/sgs/news/China-and-Higher-Education-Conference-2024>
2. <https://www.kaggle.com/competitions/predict-student-performance-from-game-play>
3. Data mining lecture
4. <https://github.com/RockyRori/avengers/tree/main/proposal>