**项目提案：通过游戏玩法预测学生表现**

1. 背景信息

利用游戏进行学习是一种创新的教育方法，将教育内容融入游戏框架中，使学习变得有趣和具有互动性。本项目利用在线教育游戏生成的时间序列数据，预测学生在游戏中的表现，以提升教育游戏的设计和有效性。本提案概述了一种结构化的方法，利用数据挖掘技术通过教育游戏增强学生的学习体验。

2. 项目目标

开发一个预测模型，在游戏的特定关卡实时预测学生是否能正确回答问题。通过此模型，帮助教育游戏开发者优化学习体验，并为教育者提供分析工具。

3. 潜在数据来源

数据集描述

- 来源：Kaggle竞赛数据集

- 内容：

- 文件：

- train.csv：训练集

- train\_labels.csv：训练集的正确答案

- test.csv：测试集

- 特征：

- session\_id，elapsed\_time，event\_name，level等

- 点击和文本信息的各种坐标

数据格式

- 每个会话有18个问题，预测每个问题的正确性。

- 数据按关卡组（0-4，5-12，13-22）顺序提供。

4. 方法描述

问题定义

使用游戏日志的时间序列数据，预测玩家在特定关卡是否能正确回答问题。

数据来源可用性

数据集在Kaggle上公开可用，可以用于分析和建模。

执行任务

1. 数据探索：

- 进行探索性数据分析以识别数据模式和关系。

- 可视化关键指标并识别重要特征。

2. 数据预处理：

- 处理缺失值并将类别变量编码为数值格式。

- 进行特征工程，创建新的预测特征。

3. 数据挖掘方法：

- 模型选择：选择适合的算法，如随机森林或梯度提升，用于分类任务。

- 训练和评估：将数据分为训练集和测试集，使用准确率和F1分数等指标评估性能。

- 超参数调整：使用网格搜索或随机搜索优化模型参数。

数据预处理、可视化和数据挖掘的理由

- 预处理：确保数据质量并为有效建模做好准备。

- 可视化：帮助理解数据模式并指导特征工程。

- 数据挖掘：使用高级算法进行准确预测，提高基于游戏的学习的教育成果。